Copertoni di plastica impermeabile proteggono l'area che si trova attorno all'impianto della Diamond Alkali Co. a Newark, nel New Jersey (Stati Uniti): infatti il suolo potrebbe venir contaminato da una diossina clorurata, la 2,3,7,8-tetraclorodibenzo-*p*-diossina, generalmente indicata con la sigla TCDD. La TCDD è un sottoprodotto della lavorazione dell'erbicida 2,4,5-T, prodotto dalla Diamond Alkali dal 1943 al 1968 in questo impianto. Le autorità dello stato del New Jersey e quelle federali hanno provveduto a questa installazione per precauzione, poiché la TCDD è tossica per molti animali. Il composto non sembra procurare gravi effetti cronici all'uomo, anche se produce inconvenienti a breve termine.

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN
numero 212
aprile 1986
anno XIX
volume XXXVI

La diossina

Si va allentando la preoccupazione per gli effetti nocivi sulla salute umana e sull'ambiente che questa sostanza può provocare: benché tossica per certi animali, non sembra, infatti, avere conseguenze a lungo termine sull'uomo

di Fred H. Tschirley

Diossina: la parola evoca innumerevoli reazioni. Molte persone sono preoccupate che il composto possa causare avvelenamento, anche in seguito a una minima esposizione. I tossicologi, che conoscono i gravi effetti che la diossina causa sugli animali di laboratorio, ma che sono incerti se gli effetti siano altrettanto gravi sull'uomo, chiedono che vengano effettuate maggiori ricerche. Alle autorità, che devono prendere decisioni in base a questi dati contraddittori, non rimane altro che chiedersi cosa fare. Dal canto mio, in qualità di chi per molti anni si è interessato al problema della diossina, spero di fornire con questo articolo un quadro utile per un giudizio sui potenziali rischi.

In realtà, il termine diossina non è usato per indicare un preciso composto, ma una famiglia di composti chimici. Il nome si riferisce alla struttura di base: due atomi di ossigeno che uniscono due anelli benzenici. La sostituzione di atomi di idrogeno sugli anelli con atomi di cloro produce numerose diossine clorurate. Ma la diossina clorurata che ci interessa è la 2,3,7,8-tetraclorodibenzo-*p*-diossina, normalmente abbreviata con la sigla TCDD. Si tratta di un sottoprodotto della fabbricazione del triclorofenolo che si usa per la produzione di due erbicidi (il più noto dei quali è l'acido 2,4,5-triclorofenossiacetico, o 2,4,5-T, uno degli ingredienti dell'Agent Orange) e dell'esaclorofene, un agente antibatterico. La TCDD viene anche prodotta in numerosi processi di combustione.

L'eredità di Agent Orange, Seveso, Times Beach, di numerosi incidenti industriali e di altri casi di esposizione di persone a quantità significative di TCDD ha destato preoccupazioni per gli effetti dannosi sulla salute, che possono risultare da una minima esposizione alla sostanza. Ma nessuno dei numerosi studi su questo problema ha finora dimostrato che la TCDD provoca effetti cronici seri sulla specie umana. Inoltre, nessun caso di decesso è stato attribuito a essa, anche se, in numerosi casi, l'esposizione è stata elevata. In questo articolo, si vuole evidenziare il grande problema della difficoltà incontrata dalle autorità che devono esprimere pareri sulla base, da un canto, di una conoscenza scientifica incompleta e, dall'altro, della generale atmosfera di paura.

La TCDD è stata riconosciuta la prima volta nel 1957 come contaminante del 2,4,5-T, quando 31 lavoratori impiegati presso un impianto per la produzione di questo erbicida, nella Germania Occidentale, presentarono disturbi dermatologici, attualmente noti come cloracne. La cloracne è un'eruzione cutanea che assomiglia all'acne e che si chiama così perché è causata dall'esposizione a vari composti chimici organici clorurati.

La consapevolezza generale che la TCDD rappresentasse un potenziale pericolo per la salute e per l'ambiente ha cominciato a farsi strada nel 1970, quando negli Stati Uniti una sottocommissione parlamentare ha tenuto una seduta sugli effetti del 2,4,5-T sull'uomo e sull'ambiente. Tra gli altri argomenti, si è discusso di uno studio del Bionetics Research Institute, secondo cui il 2,4,5-T provoca negli animali anomalie congenite. Diverse testimonianze suggerivano che il componente teratogeno del 2,4,5-T poteva essere la TCDD. Nel 2,4,5-T analizzato dal Bionetics, la diossina si trovava, come contaminante, a una concentrazione estremamente elevata: 27 ± 8 parti per milione.

Da allora sono state raccolte numerose informazioni sulle fonti della TCDD, il suo comportamento nell'ambiente e gli effetti tossici. Ben nota è la sua formazione nel corso della produzione di 2,4,5-T. La quantità di TCDD prodotta aumenta con l'aumentare della temperatura della reazione e del *p*H (grado di alcalinità).

Nel 1977 alcuni ricercatori olandesi riferirono che dibenzo-*p*-diossine policlorurate (PCDD) erano presenti nelle ceneri sospese nell'aria, provenienti da un inceneritore municipale. Normalmente, in tale inceneritore bruciano - tra le altre cose - rifiuti organici contenenti cloro. Notizie analoghe sono sopraggiunte presto dalla Svizzera, dal Canada e dal Giappone. Si pensò che quei composti si formassero in seguito alla condensazione di clorofenoli. Più tardi, alcuni dati quantitativi ricavati da un'ampia varietà di fonti di combustione hanno mostrato la presenza di TCDD che non poteva essere spiegata sulla base di preesistenti fenoli policlorurati.

Questi risultati hanno indotto R. R. Bumb della Dow Chemical Company e 12 suoi collaboratori ad avanzare, nel 1980, l'ipotesi che le PCDD potessero derivare da reazioni chimiche che avevano luogo nel fuoco in piccolissime quantità. Questa ipotesi è stata poi messa in dubbio perché le reazioni non sono mai state definite. Ciononostante, le PCDD sono state poi trovate

AMBIENTE E INQUINAMENTI

LE SCIENZE *edizione italiana di* SCIENTIFIC AMERICAN

ha dedicato all'argomento diversi articoli:

Il confinamento delle scorie radioattive
di B. L. Cohen (n. 110)

Il problema dell'anidride carbonica
di G. M. Woodwell (n. 115)

Inquinamento da PCB
di H. Suzuki (n. 116)

Dispersione di inquinanti in atmosfera a scala locale
di D. Anfossi, F. Bossa e R. Richiardone (n. 129)

Pioggia acida
di G.E. Likens, R.F. Wright, J.N. Galloway e T.J. Butler (n. 136)

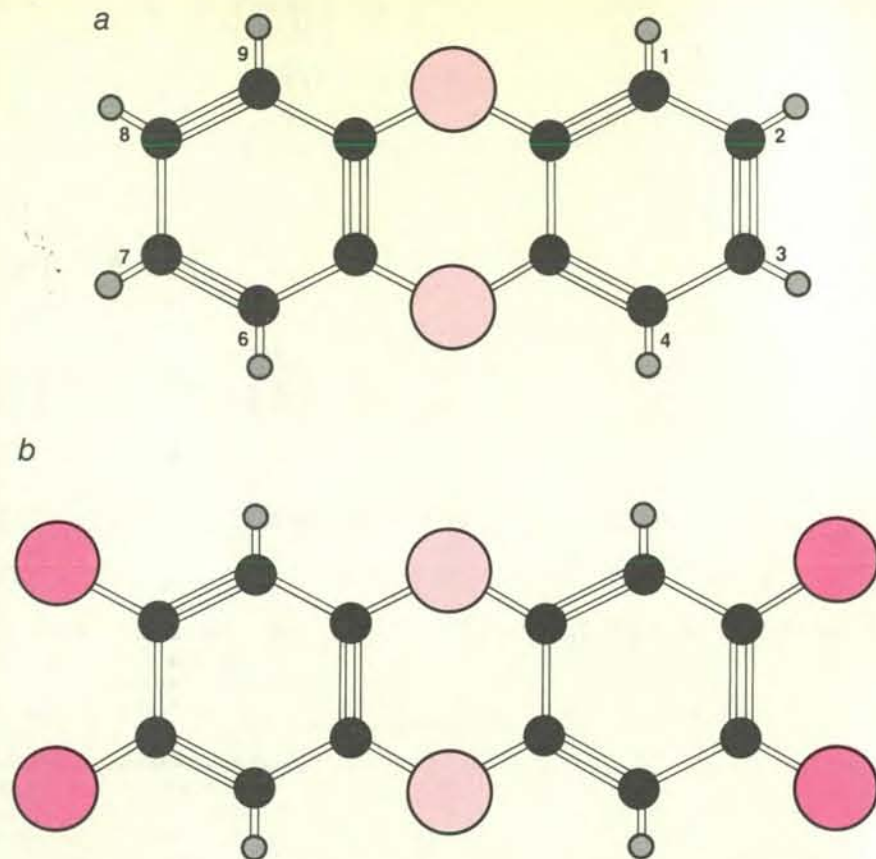
Contaminazione ambientale da fonti energetiche
di M. Dall'Aglio (n. 153)

Emissioni accidentali di radioattività
di S.A. Fetter e K. Tsipis (n. 154)

Effetti biologici di radiazioni ionizzanti di basso livello
di A.C. Upton (n. 164)

I licheni: indicatori fisiologici della qualità dell'aria
di M. Spampani (n. 167)

Anidride carbonica e clima
di B. Revelle (n. 170)

Gli effetti sul clima di una guerra nucleare
di R.P. Turco, O.B. Toon, T.P. Ackerman, J.B. Pollack e C. Sagan (n. 194)

La struttura chimica di una delle molte diossine e della TCDD. La molecola della dibenzo-*p*-diossina (*a*) consiste di due anelli benzenici legati da due atomi di ossigeno (*in colore chiaro*). Gli atomi di idrogeno (*in grigio*) uniti agli atomi di carbonio (*in nero*) nelle posizioni numerate possono essere sostituiti da uno a otto atomi di cloro formando le diossine clorurate che sono complessivamente 75. Il composto noto come TCDD (*b*) è illustrato con i quattro atomi di cloro (*in colore intenso*).

negli effluenti e nelle ceneri di così tanti processi di combustione che non esiste più il problema se si siano o no formate durante la combustione, anche se la natura precisa del processo rimane oscura.

La TCDD è stata, inoltre, identificata nel suolo e nella polvere di molti luoghi, nella fuliggine di camini di fornaci a legna, come residuo in pesci di fiumi (alcuni provenienti anche da fiumi nel cui bacino idrico non si effettuano, in realtà, operazioni industriali capaci di formare TCDD), come residuo nelle uova di gabbiani reali e, di recente, nel tessuto adiposo di più di 100 soggetti in Canada, Stati Uniti e Vietnam. Sembra che la TCDD sia, in effetti, un composto chimico ubiquitario, frequente soprattutto nei paesi industrializzati.

Ci si può chiedere allora per quale ragione non sia stata scoperta prima. Da una parte, prima degli anni settanta, non era stata ricercata con metodo. A quel tempo, in ogni caso, non la si sarebbe trovata, eccetto che in circostanze straordinarie, poiché i chimici analitici erano in grado di rivelare soltanto concentrazioni di alcune parti per milione. Da allora gli strumenti di rivelazione sono migliorati almeno un milione di volte, in modo che oggi possano essere identificate di routine concentrazioni di alcune parti per bilione. Oltre a ciò la diluizione e la distruzione della TCDD per effetto della luce possono abbassare le concentrazioni a livelli non più rilevabili. Con il miglioramento delle tecniche di analisi si potrebbe trovare TCDD in molti altri siti per ora sconosciuti.

Nei luoghi in cui è al riparo dalla luce, la TCDD è un composto estremamente resistente. Agli inizi degli anni settanta si pensava che il suo periodo di dimezzamento (ossia il tempo necessario perché la metà della quantità considerata inizialmente venisse degradata) fosse circa un anno; studi successivi, condotti negli Stati Uniti, hanno suggerito che il periodo di dimezzamento potesse avere anche la durata di tre anni. Recenti osservazioni in Italia prospettano la possibilità che, nel suolo, sia di dieci anni o anche più.

Pur in mancanza di dati precisi, si sa che la TCDD è fortemente trattenuta dalla maggior parte dei terreni. La forza di questo legame è desunta dal fatto che concentrazioni note di TCDD, applicate al suolo, sono rimaste in prossimità della superficie. Anche in terreni sabbiosi, esaminati in Florida 10-12 anni dopo l'applicazione della TCDD, la sua concentrazione nei 15 centimetri più superficiali del suolo era di 1500 nanogrammi per chilogrammo.

Nelle zone adiacenti a Seveso, in cui il grado di contaminazione è molto più elevato, la TCDD è stata trovata a una profondità di 136 centimetri. Le concentrazioni nel terreno ben al di sotto della superficie erano, inoltre, di poco più elevate, ma in modo significativo, nel 1977 rispetto al 1976, poco dopo l'incidente. La presenza di spaccature nel suolo non spiega in modo adeguato questa inaspettata distribuzione verticale.

I processi di degradazione della TCDD nel suolo sono poco conosciuti. I microrganismi degradano sì la sostanza, ma molto lentamente. S. D. Aust della Michigan State University ha scoperto un fungo (la muffa bianca *Phanerochaete chryosporium*) che provoca la decomposizione del legno e che scinde la TCDD senza che si possa osservare un effetto letale sull'organismo. La velocità di degradazione è bassa, ma si può comunque pensare di inoculare la muffa nei suoli contaminati per accelerare il processo di degradazione.

La luce del sole degrada rapidamente la TCDD mediante distacco degli atomi di cloro. La reazione richiede un donatore di idrogeno che è generalmente disponibile nell'acqua o nel rivestimento ceroso delle foglie. Esperimenti condotti da Donald G. Crosby dell'Università della California a Davis hanno dimostrato che, dopo sei ore di esposizione alla luce del sole, rimaneva inalterato il 40 per cento della TCDD stesa in strato sottile su una lastra di vetro; questa percentuale si riduceva dal 25 per cento a quasi niente quando il materiale veniva deposto in forma di gocce sulle foglie dell'albero della gomma, mentre su suolo argilloso essa era dell'85 per cento.

La TCDD si è rivelata un composto chimico estremamente tossico per gli animali. Generalmente, il primo test di tossicità eseguito sugli animali è una determinazione della LD_{50}, ossia della dose che uccide metà della popolazione presa in esame. Tra il 1973 e il 1978 la LD_{50} della TCDD è stata determinata per otto specie animali. La specie di gran lunga più sensibile fra quelle sperimentate è stata la cavia: la LD_{50} per una dose orale è stata di 0,6 milligrammi per chilogrammo di peso corporeo. Il criceto è stato l'animale meno sensibile fra quelli presi in esame: la sua LD_{50} per via orale era circa 1900 volte più alta di quella della cavia. Il suo livello intraperitoneale, tre milligrammi per chilogrammo, era circa 5000 volte più elevato della dose orale usata per la cavia.

Le ragioni della portata elevatissima della tossicità acuta (cioè a breve termine) si possono mettere in relazione con la velocità relativa a cui il corpo elimina il composto tossico: nel criceto metà della dose viene eliminata entro 15 giorni, mentre per altri animali presi in esame occorrono 30 giorni. Il criceto - come detto sopra - è molto meno sensibile della cavia; tuttavia, una LD_{50} intraperitoneale di soli tre milligrammi per chilogrammo per la TCDD indica una sostanza estremamente tossica, la cui tossicità è paragonabile a quella dell'insetticida parathion.

Molti sintomi di tipo essenzialmente acuto sono stati osservati negli esseri umani: cloracne, disturbi digestivi, effetti su alcuni sistemi enzimatici essenziali, dolori muscolari e alle giunture, effetti sul sistema nervoso e sulla psiche. Essi sono stati tuttavia transitori a eccezione di alcuni casi gravi di cloracne.

Prove supplementari hanno misurato gli effetti cronici, o a lungo termine, della TCDD su roditori, conigli e primati non

L'esigenza di una protezione globale

La diossina ha cominciato male la sua carriera: negli anni sessanta, nel Vietnam, in un contesto di violente polemiche scatenate in seguito all'impiego dei defoglianti, da alcuni assimilato a una vera e propria guerra chimica. Fatta oggetto di pesanti accuse (malformazioni congenite, tumori epatici, eccetera), la diossina aveva già acquisito una sinistra reputazione, quando accadde l'incidente di Seveso. Così, quello che sarebbe potuto essere semplicemente un caso diverso (e, in effetti, incidenti analoghi avevano già avuto luogo a molteplici riprese) divenne il simbolo della «catastrofe ecologica» con una ripercussione sul grosso pubblico nettamente superiore a quella prodotta da altri eventi, per esempio il disastro di Bhopal, in India, o l'avvelenamento da oli contaminati che fece in Spagna parecchie centinaia di vittime.

Dopo diverse altre «peregrinazioni», la diossina è venuta di nuovo alla ribalta nel gennaio 1985 a Reims, in Francia, per l'esplosione di un trasformatore contenente policlorobifenili (PCB), composti dielettrici non infiammabili, che però possono liberare, per pirolisi, policlorodibenzodiossine (PCDD) e policlorodibenzofurani (PCDF). La tetraclorodibenzodiossina (TCDD) è soltanto un isomero particolare delle 135 PCDD e degli 85 PCDF, che possono teoricamente formarsi in circostanze del genere. La situazione a Reims era nuova e inquietante per almeno due ragioni: in primo luogo, contrariamente a ciò che era accaduto a Seveso, dove - secondo numerosi osservatori - la reazione delle autorità responsabili era stata tempestiva, a Reims vi fu un grave ritardo, durante il quale numerose persone (circa 300) rimasero esposte in grado diverso a PCDD e a PCDF. Inoltre, mentre la chimica del triclorofenolo ha un campo d'azione relativamente ristretto, trasformatori e condensatori contenenti policlorobifenili sono installati a migliaia nei pressi di zone residenziali o industriali, dove di tanto in tanto si segnalano incidenti.

Come viene egregiamente spiegato in queste pagine da Fred H. Tschirley, esiste un notevole salto tra i dati sperimentali (parecchie migliaia di pubblicazioni sono state dedicate alla diossina) e i dati relativi alla specie umana, molto più rari e nell'insieme rassicuranti, anche se la cloracne può essere considerata un'affezione grave a causa della sua cronicità e dei postumi antiestetici che ne derivano. Sussiste, inoltre, un dubbio sul rischio di tumore, suscettibile di comparire molto tempo dopo l'esposizione.

In casi come questi il clinico ha un compito difficile: valutare un rischio estrapolando all'uomo dati sperimentali. È indubbio che certe informazioni, apparentemente molto inquietanti, debbano essere interpretate. Per esempio, è vero che la diossina ha una dose letale al 50 per cento (LD_{50}) che è inferiore a un milligrammo per chilogrammo di peso corporeo, ma è anche vero che la digitalina, un farmaco di cui si fa grande uso, può uccidere un uomo alla dose di soli 100 microgrammi per chilogrammo di peso corporeo. Certamente, l'incenerimento di scorie contenenti policlorobifenili può liberare diossina, ma è anche vero che un semplice fuoco di legna e numerose altre combustioni liberano idrocarburi poliaromatici, che sono nettamente più cancerogeni. Beninteso, tutto ciò non vuol dire affatto che si possono trascurare i rischi legati alla diossina; anzi indica la necessità improrogabile di prevedere una protezione globale della salute e dell'ambiente umani, evitando di privilegiare certi rischi per ragioni del tutto irrazionali.

Tuttavia, nell'incidente di cui io e il mio gruppo ci siamo dovuti occupare, l'inquietudine nasceva piuttosto dagli altri PCDD e PCDF, per i quali si avevano dati molto più frammentari. In effetti, per la TCDD si può arrivare oggi a una valutazione relativamente precisa del rischio grazie a un numero sufficiente di dati relativi all'uomo. Nel caso in cui, però, siano chiamati contemporaneamente in causa molteplici composti con strutture chimiche simili, ma con effetti tossici molto diversi, l'analisi tossicologica e la valutazione del rischio si complicano in modo singolare.

A più di un anno di distanza, l'incidente di Reims non sembra aver provocato - almeno fino a oggi - conseguenze più gravi di quello di Seveso. Ma ogni medico sa bene che è spesso più difficile promettere, a chi è in buona salute, di mantenerlo sano che non a un malato di guarirlo.

Sylvain Dally
Clinique toxicologique
Hôpital Fernand-Widal
Parigi

Sono qui riportati i passaggi che conducono alla formazione della TCDD durante la produzione di 2,4,5-triclorofenolo (TCP). Questo composto, sotto forma di triclorofenato (*2*), viene usato per produrre gli erbicidi 2,4,5-T e Silvex, nonché l'agente antibatterico esaclorofene. Nella produzione di TCP, una molecola di 1,2,4,5-tetraclorobenzene viene idrolizzata (*1*) in presenza di soda caustica e forma il 2,4,5-triclorofenato sodico (*2*). Due molecole di questo (*3*) si combinano e formano la TCDD (*4*) con perdita di atomi di sodio (*in bianco*) e di cloro. Quanto più elevate sono la temperatura e l'alcalinità, tanto maggiore è la quantità di TCDD che si forma. Se la temperatura supera i 180 gradi centigradi, la reazione è esotermica: questa è stata la causa degli incidenti.

umani. La cloracne, che è l'indicatore più sensibile dell'esposizione di esseri umani alla TCDD, si è manifestata anche nei conigli, nei primati non umani e nei topi glabri. La pelle può diventare secca e squamarsi. Alcune specie perdono il pelo e alcuni primati non umani perdono le unghie, apparentemente senza dolore.

La TCDD causa anche effetti sulla riproduzione degli animali di laboratorio. Palatoschisi e anomalie a carico del rene sono state causate nei discendenti di topi esposti a una dose giornaliera variabile da uno a tre nanogrammi per chilogrammo di peso corporeo. Dosi analoghe hanno causato la morte nei feti di ratti. Nelle scimmie, una dose giornaliera di 1,7 nanogrammi per chilogrammo, per due anni, ha causato aborti in quattro gravidanze su sette.

Nei ratti e nei topi la TCDD è un cancerogeno collaudato: pur essendovi una certa variabilità tra i risultati delle analisi, le prove eseguite da ricercatori diversi concordano abbastanza. Il fegato subisce le principali conseguenze, sia nei ratti sia nei topi, ma in alcuni studi sono stati presi in considerazione anche il cervello, il sistema respiratorio e la tiroide. È importante riconoscere che un referto oncogeno è stato riportato solo dopo che l'animale aveva ingerito dosi elevate di TCDD per un lungo periodo di tempo. La sostanza non si è rivelata oncogena nei ratti a una dose giornaliera di 1-1,4 nanogrammi per chilogrammo di peso corporeo, mentre nei topi essa non lo è stata nemmeno a una dose compresa tra uno e 30 nanogrammi. Uno studio condotto da R. J. Kociba della Dow Chemical e collaboratori ha inoltre dimostrato che i ratti tolleravano una dose giornaliera di un nanogrammo per chilogrammo di peso corporeo per un periodo di due anni, senza che in essi si manifestassero effetti tossici.

I risultati delle prove sugli animali sono, tuttavia, serviti a far crescere l'interesse sugli effetti della TCDD nell'uomo. In numerose occasioni soggetti umani sono stati esposti a dosi «elevate» di TCDD. In questo contesto, «elevato» è un termine relativo poiché, con una eccezione (un gruppo di prigionieri offertosi volontario per sottoporsi alle prove con la TCDD) non è nota con precisione la quantità di sostanza alla quale ciascuno dei soggetti è stato esposto. Il criterio da me impiegato per distinguere l'esposizione elevata dalla bassa è quello della cloracne, che può conseguire o no all'esposizione.

Il caso dei prigionieri è importante perché si conoscono le quantità di TCDD applicate alla loro pelle. Nel primo esperimento 60 volontari sono stati trattati con concentrazioni che variavano da 200 a 8000 nanogrammi (da tre a 114 nanogrammi per chilogrammo di peso corporeo, per una persona che pesava 70 chilogrammi), e la somministrazione è stata ripetuta due settimane più tardi. Le dosi scelte sono state quelle che avevano causato cloracne quando erano state applicate alle orecchie dei conigli. Nessuno dei volontari ha accusato la comparsa di cloracne e non sono stati neppure osservati altri sintomi. Il secondo esperimento ha interessato 10 prigionieri volontari che sono stati trattati con 107 000 nanogrammi di TCDD per chilogrammo: in otto di essi si è manifestata cloracne, ma non sono stati notati altri sintomi.

Da questi esperimenti si può concludere solo che la TCDD causa sì cloracne negli uomini, quando la dose somministrata è sufficientemente elevata, ma che gli uomini sono meno sensibili dei conigli. Le prove non hanno tuttavia permesso di identificare una soglia per lo sviluppo della cloracne nei soggetti umani: un'informazione che sarebbe di grande importanza.

Il numero dei soggetti umani che sono stati esposti a dosi di TCDD non può essere determinato con precisione, ma deve essere nell'ordine delle migliaia. Alistair Hay dell'Università di Leeds ha stimato che solo nell'industria chimica circa 2000 lavoratori hanno subito un'esposizione elevata. A bassi livelli sono stati senza dubbio esposti coloro che hanno manipolato gli erbicidi 2,4,5-T e Silvex, in cui la TCDD era un agente contaminante; i veterani del Vietnam, esposti all'Agent Orange (costituito per il 50 per cento da 2,4,5-T); gli abitanti di Times Beach, nel Missouri, dove scarichi oleosi contenenti TCDD sono stati riversati in molti luoghi sul terreno, gli operai di industrie chimiche che producono materiali contenenti il composto e numerose migliaia di individui che hanno consumato alimenti (in modo particolare pesce) contenenti tracce di TCDD, oppure sono stati esposti alla caduta di ceneri provenienti da processi di combustione che formano TCDD. Il numero totale di individui sottoposto a tali basse esposizioni probabilmente ammonta a milioni.

La possibilità di effetti cronici conseguenti all'esposizione alla TCDD provoca nell'opinione pubblica un'attenzione maggiore di quella per gli effetti acuti. Un aspetto di questo problema, del quale poco si conosce, è l'effetto provocato da un'esposizione prolungata a bassi livelli della sostanza, quale si potrebbe avere in un ambiente professionale, o da esposizioni accidentali, per esempio alla caduta di ceneri di combustione o a pesce che contiene bassi livelli di TCDD

Uno sguardo ad alcuni dei principali casi di esposizione alla TCDD, in ordine approssimativo di gravità, rivela pochi effetti cronici inequivocabili, seppure ve ne siano. L'incidente di Seveso, nel 1976, espose circa 37 000 persone di tutte le età a considerevoli quantità di TCDD. Un numero relativamente piccolo di queste persone ha manifestato effetti transitori, come cloracne (184 casi, di cui 164 erano bambini al di sotto di 15 anni di età), cefalee e disturbi digestivi, mentre non sono stati identificati casi di effetti a lungo termine, come anomalie congenite e danni cromosomici. È ancora troppo presto per dire se l'incidenza del cancro in questi soggetti si discosti dai valori normali.

SPECIE	VIA	LD_{50} (microgrammi per chilogrammo)
CAVIA (MASCHIO)	ORALE	0,6
CAVIA (FEMMINA)	ORALE	2,1
CONIGLIO (MASCHIO, FEMMINA)	ORALE	115
CONIGLIO (MASCHIO, FEMMINA)	CUTANEA	275
CONIGLIO (MASCHIO, FEMMINA)	INTRAPERITONEALE	252-500
SCIMMIA (FEMMINA)	ORALE	< 70
RATTO (MASCHIO)	ORALE	22
RATTO (FEMMINA)	ORALE	45-500
TOPO (MASCHIO)	ORALE	<150
TOPO (MASCHIO)	INTRAPERITONEALE	120
CANE (MASCHIO)	ORALE	30-300
CANE (FEMMINA)	ORALE	>100
RANA	ORALE	1000
CRICETO (MASCHIO, FEMMINA)	ORALE	1157
CRICETO (MASCHIO, FEMMINA)	INTRAPERITONEALE	3000

La tossicità della TCDD sugli animali viene accertata mediante la LD_{50} cioè la dose (in microgrammi per chilogrammo di peso corporeo) che uccide la metà degli individui del gruppo esaminato.

Un incidente in un impianto della Monsanto a Nitro, nel West Virginia, avvenuto nel 1949, provocò un'esposizione di più di 200 operai alla TCDD. Dei 122 operai ai quali si manifestò la cloracne, 121 furono tenuti sotto controllo per i successivi 30 anni. Il numero complessivo di decessi in quel gruppo non si allontanò in modo significativo dal numero prevedibile per la popolazione in generale; non aumentò neppure il numero di decessi dovuti al cancro o a disturbi dell'apparato circolatorio. Risultati analoghi si ebbero come conseguenza di altri incidenti industriali, a eccezione di due in cui, in piccoli gruppi di persone esposte alla TCDD, fu rilevato un maggior numero di decessi per cancro.

Un tipo particolare di cancro (sarcoma dei tessuti molli, che genericamente indica più di 100 tipi diversi di cancro con rara incidenza) ha focalizzato gli interessi a causa di un'indagine condotta da Lennart Hardell dell'Università di Umea su lavoratori forestali svedesi. Hardell ha concluso che l'esposizione al 2,4,5-T (e pertanto alla TCDD) aveva causato la comparsa di sarcomi dei tessuti molli con un'incidenza sei volte superiore alla norma.

Questo studio ha indotto a indagare sugli addetti a impianti chimici negli Stati Uniti, che erano stati esposti al 2,4,5-T e ad altri prodotti chimici. Sono stati così scoperti sette casi evidenti di sarcoma dei tessuti molli, il che ha sostanzialmente elevato il livello di interesse per questo argomento. Eventi successivi hanno messo in rilievo le difficoltà di una diagnosi accurata di quel sarcoma e di una identificazione accurata degli individui esposti. A una conferenza nel 1983, Marilyn A. Fingerhut del National Institute of Occupational Safety and Health riferì che, di fatto, due delle sette persone erano morte per forme di cancro diverse dal sarcoma dei tessuti molli. Inoltre, l'esposizione alla TCDD di altre tre persone non poteva essere ben documentata. Tali risultati sono ben lontani dal riuscire a rappresentare una seria testimonianza che la TCDD sia causa di sarcoma dei tessuti molli.

Anche altri studi non riescono ad avvalorare l'ipotesi di Hardell. Nello stato di Washington, tra le professioni in cui i lavoratori sarebbero stati esposti alla TCDD, non si è trovato alcun quadro coerente di decessi causati da sarcoma dei tessuti molli. Uno studio condotto in Finlandia non ha rilevato casi del genere tra i 1900 individui che avevano maneggiato erbicidi; neppure il tasso di mortalità dovuta a cause naturali era diverso da quello della totalità della popolazione maschile in Finlandia. La US Air Force nel suo studio, effettuato a Ranch Hand su circa 1200 addetti militari che avevano partecipato alle operazioni di nebulizzazione di Agent Orange in Vietnam, non ha trovato casi di sarcoma dei tessuti molli. Infine attente osservazioni eseguite dalla Veterans Administration su 85 000 veterani, autoselezionatisi, hanno indicato un minor numero di casi di sarcomi dei tessuti molli, rispetto alla media nazionale.

Sono oggetto di indagine anche gli effetti sulla riproduzione, in conseguenza delle scoperte fatte sugli animali. Il caso più famoso che sostiene tali effetti nell'uomo, è comunemente noto come studio Alsea II, ed è stato compiuto dalla US Environmental Protection Agency. In esso si stabilisce un nesso tra le nebulizzazioni di 2,4,5-T sulle foglie e l'aborto spontaneo in donne gravide ad Alsea, nell'Oregon.

L'Alsea II ha suscitato molte critiche, soprattutto da parte di un gruppo interdisciplinare della Oregon State University. Questo gruppo concludeva che, in base ai dati rilevati dall'ente non era possibile dimostrare una connessione tra impiego dell'erbicida e aborto spontaneo. Altri studi condotti in Australia, Ungheria, Nuova Zelanda e Stati Uniti non sono riusciti a trovare un legame tra l'uso del 2,4,5-T e anomalie congenite.

A causa dell'estrema tossicità acuta della TCDD e dei suoi molteplici effetti cronici sugli animali, gli enti che provvedono a regolamentare la materia hanno dovuto prendere in considerazione ciò che è necessario fare per proteggere le persone dall'esposizione al composto. Tranne che in pochi casi, si devono estrapolare i dati riferiti agli animali e li si devono adattare agli esseri umani, nonostante che non sia stata ancora accertata la validità di questo tipo di estrapolazione. La mancanza di un metodo semplice e accurato per determinare se la TCDD si trova nei tessuti di individui che sono stati esposti al composto e a che livello fa aumentare le difficoltà. (La prova richiede un intervento chirurgico per ottenere campioni di fegato e di tessuto adiposo in cui la TCDD si fissa.) Senza questa informazione non può essere stabilita una relazione tra dose e risposta.

La TCDD è stata definita il composto chimico sintetico più tossico che l'uomo conosca. Se la sua tossicità acuta nei confronti della cavia, e perfino del ratto e del topo, ne è il parametro, l'affermazione è probabilmente corretta. Se, invece, il parametro fosse determinato dalla minor tossicità per il criceto, l'affermazione sarebbe sicuramente non corretta. Non c'è comunque bisogno di arzigogolare: la TCDD è incontestabilmente un composto chimico dotato di estrema tossicità per gli animali di laboratorio. Inoltre, in seguito alla somministrazione di basse dosi sono stati dimostrati gravi effetti cronici su cavie. È quindi comprensibile l'interesse riguardo agli effetti che la TCDD può avere anche sulla salute e sull'ambiente umano.

Quando si considerano composti tossici, un ente di controllo ha poche scelte al di là dall'estrapolare all'uomo i dati ricavati sugli animali. Per di più gli effetti sulla salute umana sono raramente dimostrabili nel caso di composti chimici sparsi nell'ambiente, ai quali la gente è variamente esposta in dosi subacute, che possono essere solo stimate (e solo con una grossolana approssimazione). Un esempio calzante è quello dell'aflatossina, il prodotto di una muffa che si sviluppa comunemente su semi oleosi immagazzinati, come le arachidi. Nelle prove su animali l'aflatossina si è dimostrata uno dei più potenti agenti cancerogeni noti, ma non è ancora stato provato che essa abbia questo effetto sugli esseri umani.

La diversità nella reazione agli stimoli è una caratteristica degli organismi viventi. Non fanno eccezione alla regola generale le reazioni alle tossine. I soggetti umani possono essere sensibili a una data tossina più, o meno, o nella stessa misura di un animale di laboratorio. L'estrapolazione non è né arte, né scienza: è semplicemente il modo più razionale per valutare un rischio in assenza di dati definitivi. Pertanto gli interventi di controllo continuano a essere basati sui dati relativi agli animali, anche quando i dati sulla nostra specie, benché non definitivi, possono essere sufficientemente convincenti da far dire, sul piano scientifico, che il rischio per l'uomo è stato valutato esageratamente.

Sembra questo il caso della TCDD. I ricercatori sono tutti d'accordo sul fatto che la TCDD sia meno tossica per l'uomo che per gli animali di laboratorio, ma le informazioni disponibili non hanno forza sufficiente da indurre un cambiamento nell'atteggiamento normativo verso una maggiore o una minore limitazione dell'esposizione al composto. Dubito che la prova diretta degli effetti della TCDD sugli esseri umani potrà mai essere più decisiva, o meno, di quanto lo sia adesso.

L'opinione pubblica nei confronti di una sostanza tossica costituisce un fattore di importanza determinante nell'atteggiamento assunto da un ente di controllo. La gente ha sentito molto parlare degli effetti acuti e cronici della TCDD sugli animali di laboratorio, ma poco della mole non indifferente di dati che dimostrano come gli esseri umani siano meno sensibili. I rapporti iniziali sulla tossicità acuta della TCDD, seguiti da altri rapporti sulla sua cancerogenità e sugli effetti che essa provoca sulla riproduzione hanno instillato nella gente una paura che probabilmente non potrà essere fugata neppure da un'informazione adeguata su un'esperienza equivalente riguardante soggetti umani. L'ente di controllo pertanto si trova nella condizione di avere a che fare non solo con i dati delle indagini disponibili, ma anche con i timori della gente.

La US Environmental Protection Agency ha risposto ai timori della gente con una serie di regolamenti intesi a tenere sotto controllo sia la formazione della TCDD sia la sua liberazione nell'ambiente, e a limitare l'esposizione dei singoli individui a essa. Tali regolamenti potrebbero essere resi più o meno rigidi sulla base di nuove verifiche. Ciò che l'ente non ha fatto - e si potrebbe dire che avrebbe la responsabilità di fare - è tentare di dissipare il timore della gente, dimostrando che l'esposizione in ambienti a bassa concentrazione di TCDD non dovrebbe avere gravi effetti cronici sui soggetti umani.

La questione della TCDD è ancora più esacerbata dal fatto che la TCDD viene messa in relazione con il programma di defogliazione nel Vietnam: un programma impopolare in una guerra impopolare. Sono stati ampiamente divulgati i numerosi e diversi effetti sulla salute che, secondo i veterani del Vietnam, sarebbero stati causati dall'esposizione all'Agent Orange. La gente è generalmente convinta che le querele erano state fatte ritirare, fuori dal tribunale, con 180 milioni di dollari e molti credono che tale composizione della controversia fosse un'ammissione di colpa da parte delle industrie chimiche che avevano prodotto l'Agent Orange. Evidentemente pochi conoscono l'affermazione fatta dal giudice federale Jack B. Weinstein ai legali dei querelanti: «in nessun caso avete dimostrato l'effettiva insorgenza sulla salute delle conseguenze citate».

Un problema che preoccupa a proposito della TCDD è l'utilizzazione appropriata delle risorse scientifiche. A. L. Young dell'Office of Science and Technology Policy ha calcolato che il governo federale avrà speso più di due miliardi di dollari per la ricerca e per altre questioni connesse con la diossina prima che siano completati tutti gli importanti studi attualmente in corso. Altro dispendio sia in termini di tempo sia di denaro è stato sostenuto da industrie chimiche, organizzazioni private ed enti governativi. L'esborso complessivo è una cifra enorme per un argomento di importanza discutibile.

Due anni fa in una conferenza sulla diossina, tenuta nel Michigan, si giunse alla conclusione che il caso della TCDD è relativamente meno importante di molti altri argomenti e che le risorse scientifiche limitate della nazione dovrebbero essere indirizzate verso altre questioni che costituiscono una minaccia maggiore. Sulla base dei dati emersi finora, quella conclusione sembra ancora valida.

DATA	LAVORATORI ESPOSTI	LUOGO DELL'INCIDENTE	NOTE
1949	250	Impianto della Monsanto a Nitro, nel West Virginia (Stati Uniti)	122 casi di cloracne accertati; 32 decessi contro i 46,4 attesi; nessun aumento dei decessi per neoplasie maligne o per disturbi circolatori
1953	75	Impianto della BASF a Ludwigshafen (Repubblica Federale Tedesca)	55 casi di cloracne, di cui 42 gravi; 17 decessi su una previsione dagli 11 ai 25 (quattro casi di carcinoma gastrointestinale e due di carcinoma polmonare); danni agli organi di senso e al fegato
1956	?	Impianto della Rhone-Poulenc a Grenoble (Francia)	17 casi di cloracne, inoltre tassi elevati di lipidi e di colesterolo nel sangue
1963	106	Impianto della NV Philips ad Amsterdam (Paesi Bassi)	44 casi di cloracne, di cui 42 gravi: 21 anche con danni interni o disturbi al sistema nervoso centrale; otto decessi (sei per probabile infarto miocardico); alcuni sintomi di affaticamento
1964	61	Impianto della Dow Chemical a Midland, nel Michigan (Stati Uniti)	49 casi di cloracne; quattro decessi contro i 7,8 attesi; tre decessi per tumore rispetto agli 1,5 attesi; un sarcoma dei tessuti molli
1965-69	78	Impianto di Spolana, nei pressi di Praga (Cecoslovacchia), dove hanno avuto luogo continue perdite	78 casi di cloracne; cinque decessi; dei 50 lavoratori che sono stati analizzati per più di 10 anni, molti mostrano ipertensione, elevato tasso di lipidi e di colesterolo nel sangue; stato prediabetico; danni gravi al fegato e al sistema nervoso
1966	?	Impianto della Rhone-Poulenc a Grenoble (Francia)	21 casi di cloracne
1968	90	Impianto della Coalite & Chemical nel Derbyshire (Regno Unito)	79 casi di cloracne; un decesso per trombosi coronarica
1976	156	Impianto della ICMESA a Seveso	Addetti all'impianto e abitanti della zona ancora sotto esame; più di 500 residenti trattati per presunti sintomi tossici; 134 casi confermati di cloracne; tasso normale di mortalità complessiva

Gli incidenti industriali hanno esposto più di 800 lavoratori a TCDD. L'incidente di Seveso del 1976 ha interessato anche 37 000 abitanti del luogo. I dati sono dell'American Medical Association.

La decomposizione della TCDD ha luogo quando i raggi ultravioletti che compongono la luce solare scindono gli atomi di cloro della molecola. Sono qui illustrati sei passaggi della reazione. Nel terreno, dove la luce solare non può penetrare, la TCDD tende a persistere anche per 10 anni.

L'astronomia islamica

Mentre attraversava un periodo di grave declino nell'Europa medievale, l'astronomia fioriva nel mondo islamico, dove la scienza degli antichi greci venne conservata per essere poi trasmessa alla cultura del Rinascimento

di Owen Gingerich

Gli storici che ricostruiscono lo sviluppo dell'astronomia dall'Antichità al Rinascimento ricorrono a volte, per designare il periodo di tempo compreso fra l'VIII e il XIV secolo, alla denominazione di «periodo islamico». Durante quell'intervallo, infatti, la maggior parte dell'attività astronomica si svolse nel Medio Oriente, nell'Africa settentrionale e nella Spagna moresca. Mentre l'Europa languiva nell'«Età buia», la fiaccola della cultura antica era passata in mani musulmane. I dotti islamici ne tennero accesa la fiamma, che trasmisero poi all'Europa del Rinascimento.

Due circostanze favorirono la crescita dell'astronomia nei paesi islamici. La prima fu la vicinanza geografica al mondo della cultura antica, unita alla tolleranza per i dotti di altre credenze religiose. Nel IX secolo la maggior parte dei testi scientifici greci fu tradotta in arabo, compresa la *Syntaxis* di Tolomeo, che era l'apice dell'astronomia antica. Proprio attraverso queste traduzioni le opere greche divennero note in seguito nell'Europa medievale. (In effetti la *Syntaxis* è nota ancora oggi come *Almagesto*, dall'arabo *al-Magistī* - corruzione del greco *e megiste* [*mathematike syntaxis*] - che significa «la massima» [raccolta matematica].)

Il secondo impulso venne dalla religione islamica, la quale presentava una numerosa serie di problemi di astronomia matematica, connessi per lo più al computo del tempo. Nella soluzione di questi problemi i dotti islamici andarono molto oltre i metodi matematici greci. Questi sviluppi, particolarmente nel campo della trigonometria, fornirono gli strumenti essenziali per la creazione dell'astronomia occidentale del Rinascimento.

Le tracce dell'astronomia medievale islamica sono manifeste ancora oggi. Quando ci riferiamo allo zenit, all'azimut o all'algebra, o quando menzioniamo le stelle Vega, Altair e Deneb, che formano un triangolo facilmente identificabile nel cielo d'estate, usiamo parole di origine araba. Eppure, benché la storia di come l'astronomia greca passò agli arabi sia relativamente ben nota, solo ora si sta cominciando a scrivere la storia della sua trasformazione da parte di dotti islamici e della successiva ritrasmissione nell'Occidente latino. Migliaia di manoscritti rimangono da esaminare. Ciononostante è possibile offrire per lo meno un abbozzo frammentario di questo processo.

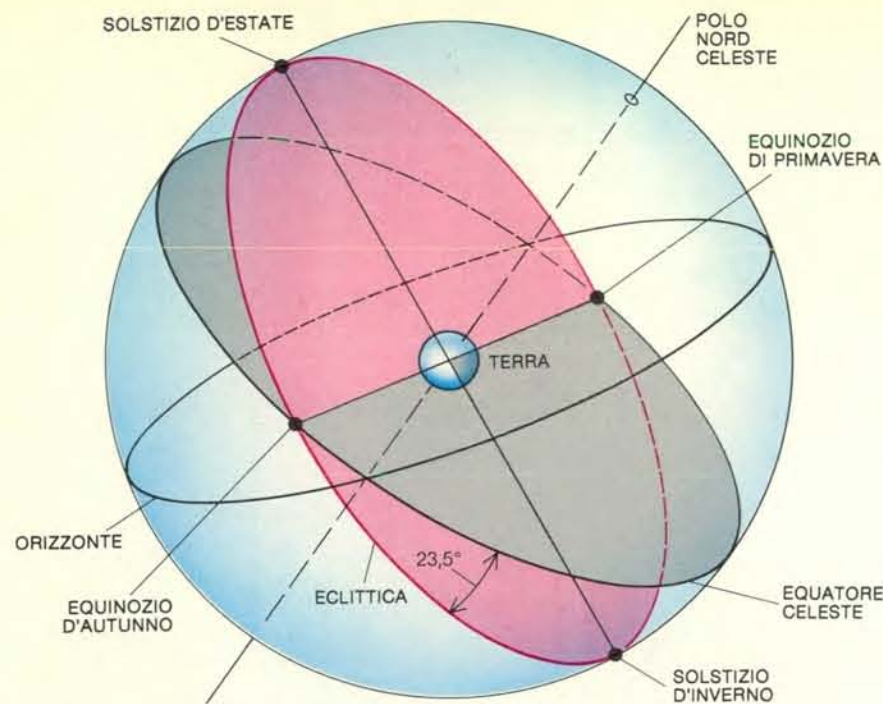
La «Casa della sapienza»

Le basi della scienza islamica in generale e dell'astronomia in particolare furono gettate due secoli dopo l'emigrazione del profeta Maometto dalla Mecca a Medina, nel 622 d.C. Questo evento, noto come Egira, segna l'inizio del calendario islamico. I primi secoli dell'Islam furono caratterizzati da un'espansione rapida e turbolenta. Solo verso la fine del II e all'inizio del III secolo del calendario dell'Egira la situazione divenne abbastanza stabile e cosmopolita da consentire la fioritura delle scienze. La nuova dinastia abbaside, che nel 750 era ascesa al califfato (assumendosi la direzione del mondo islamico) e che nel 1762 aveva fondato la capitale Bagdad, cominciò a patrocinare traduzioni di testi greci. In pochi decenni furono tradotte in arabo le principali opere scientifiche dell'Antichità, fra le quali gli scritti di Galeno, Aristotele, Euclide, Tolomeo, Archimede e Apollonio. Il lavoro di traduzione fu compiuto da dotti cristiani e pagani, oltre che musulmani.

Il patrono più generoso di questo sforzo fu il califfo al-Ma'mūn, salito al potere nell'813. Al-Ma'mūn fondò un'accademia chiamata «Casa della sapienza» (*Bait al-ḥikmah*), che affidò alla direzione di Ḥunayn ibn Isḥaq al-'Ibādī, un cristiano nestoriano che aveva una padronanza eccellente del greco. Ḥunayn divenne il più famoso fra tutti i traduttori di testi greci. Si devono a lui versioni in arabo di Platone, di Aristotele e dei loro commentatori, oltre a traduzioni delle opere dei tre padri fondatori della medicina greca, Ippocrate, Galeno e Dioscoride.

Il principale traduttore di opere matematiche e astronomiche attivo nell'accademia fu un pagano, Thābit ibn Qurra. Thābit era stato in origine un cambiavalute sulla piazza del mercato di Harran, una cittadina del nord della Mesopotamia che era il centro di un culto astrale. Thābit sostenne coraggiosamente che erano stati gli adepti di quel culto a colonizzare inizialmente il paese, a fondare città, a costruire porti e a scoprire la scienza, ma, nonostante queste opinioni eterodosse, egli venne tollerato nella capitale islamica, dove scrisse più di cento trattati scientifici, fra i quali un commento all'*Almagesto* di Tolomeo.

Un altro astronomo matematico affiliato alla Casa della sapienza fu al-Khwārizmī, la cui *Algebra*, dedicata ad al-Ma'mūn, potrebbe essere stata il primo libro sull'argomento in arabo. Pur non essendo un'opera particolarmente degna di nota sul piano scientifico, ha il merito di aver contribuito a introdurre nel mondo islamico metodi indù oltre che greci. Un po' dopo il 1100 essa fu tradotta in latino dall'inglese Roberto di Chester, che si era recato in Spagna a studiare matematica. La traduzione, che si apre con le parole *Dicit Algoritmi* («Dice Algoritmi», da cui la parola moderna «algoritmo»), esercitò una forte influenza sull'algebra occidentale del Rinascimento.

La carta stellare della costellazione di Perseo è una copia islamica medievale di un disegno eseguito nel X secolo dall'astronomo persiano 'Abd al-Raḥmān al-Ṣūfī. Al-Ṣūfī fece una revisione del catalogo stellare compilato nel II secolo d.C. da Tolomeo, il principale astronomo dell'Antichità. L'opera di Tolomeo, la *Syntaxis*, che cataloga oltre 1000 stelle, è nota ancora oggi come *Almagesto*, dall'arabo *al-Magistī*. La fotografia, di Owen Gingerich, riproduce la pagina di un codice ora appartenente alla Biblioteca nazionale egiziana del Cairo.

Gli astronomi antichi pensavano che fosse la rotazione della sfera celeste attorno alla Terra a spiegare il moto apparente diurno del Sole e delle stelle. In base a questa concezione, i poli e l'equatore della sfera celeste sono proiezioni in cielo dei poli e dell'equatore della Terra. In conseguenza della rotazione diurna, da est a ovest, della sfera attorno al suo asse, il Sole e le stelle si muovono lungo cerchi concentrici attorno ai poli. Il Sole si muove un po' più lentamente delle stelle, cosicché nel corso dell'anno percorre sulla sfera celeste una traiettoria circolare, da ovest a est, detta eclittica. Il corso del Sole (l'eclittica) interseca l'equatore celeste in due punti, corrispondenti agli equinozi; al solstizio d'estate e al solstizio di inverno il Sole si trova alla sua massima distanza dall'equatore, circa 23,5 gradi verso nord o verso sud. Un osservatore che sia situato sulla Terra vede solamente la metà della sfera celeste che si trova al di sopra dell'orizzonte.

La sua influenza è inoltre ancora avvertibile nell'intera matematica e nella scienza in generale: essa segnò l'introduzione in Europa dei «numerali arabi». Assieme a certi procedimenti trigonometrici, gli arabi avevano preso a prestito dall'India un sistema di numeri comprendente lo zero. I numerali indiani esistettero nel mondo islamico in due forme, e fu la forma occidentale a passare, attraverso la Spagna, nell'Europa medievale. Questi numerali, con un segno esplicito per lo zero, sono molto più efficienti, per il calcolo, dei numerali romani.

Un altro astronomo, attivo a Bagdad nel IX secolo, fu Aḥmad al-Farghānī (l'Alfraganus dei latini). La sua opera astronomica più importante furono i *Rudimenta astronomica* o *Elementi di astronomia* (*Jawāmi'*), che contribuirono alla diffusione delle parti più elementari e non matematiche dell'astronomia geocentrica di Tolomeo.

I *Rudimenta* ebbero un'influenza considerevole in Occidente. Vennero tradotti due volte in latino a Toledo, una volta da Giovanni da Siviglia (Johannes Hispalensis) nella prima metà del XII secolo e una seconda volta, in modo molto più completo, da Gherardo da Cremona alcuni decenni dopo.

La traduzione di Gherardo dell'opera di Alfraganus fu per Dante la principale fonte di conoscenza dell'astronomia tolemaica. (Nel *Paradiso* il poeta sale verso l'empireo passando per i cieli dei pianeti, che erano sfere incentrate sulla Terra.) In Occidente la versione più nota dei *Rudimenta* fu però quella anteriore di Giovanni da Siviglia. Essa servì di base alla *Sphaera* di Sacrobosco, esposizione ancor più semplificata dell'astronomia sferica, scritta all'inizio del XIII secolo da John of Holywood (Johannes de Sacrobosco). Nelle università del mondo occidentale la *Sphaera* di Sacrobosco rimase il libro di testo tradizionale per secoli. Nell'epoca della stampa ebbe più di 200 edizioni, prima di essere soppiantato da altri testi all'inizio del Seicento. Con l'eccezione degli *Elementi* di Euclide nessun altro testo scientifico ebbe una fortuna così costante e duratura.

Così dalla Casa della sapienza nell'antica Bagdad, con la sua tolleranza e la sua miscela unica di culture, uscì una sequenza impressionante non solo di traduzioni di opere scientifiche e filosofiche greche, ma anche di commenti e di trattati originali. Nel 900 erano ormai state poste le basi per la grande fioritura di una scienza internazionale, con un'unica lingua - l'arabo - come proprio veicolo.

Religione e astronomia

Un forte contributo alla fioritura dell'astronomia nel mondo islamico venne dalla religione, che si trovò a dover affrontare una varietà di problemi di astronomia matematica, e specificamente di geometria sferica.

Al tempo di Maometto tanto i cristiani quanto gli ebrei osservavano giorni sacri, come la Pasqua, la cui data era determinata dalle fasi della Luna. Entrambe le comunità avevano dovuto confrontarsi con il fatto che i mesi lunari, di (circa) 29,5 giorni, non sono commensurabili con l'anno solare di 365 giorni: la somma di 12 mesi lunari dà infatti solo 354 giorni. Per risolvere il problema, cristiani ed ebrei avevano adottato uno schema fondato su una scoperta fatta attorno al 430 a.C. dall'astronomo ateniese Metone. Nel ciclo metonico di 19 anni c'erano 12 anni di 12 mesi lunari e 7 anni di 13 mesi lunari. L'inserimento periodico di un tredicesimo mese manteneva le date del calendario in accordo con le stagioni.

A quanto pare, però, nell'aggiunta del mese intercalare non si seguì sempre una regola precisa; sovrani di pochi scrupoli aggiungevano a volte il tredicesimo mese quando più conveniva loro. Per Maometto il mese intercalare era opera del demonio. Nel *Corano* (cap. 9, versetto 36) egli decretò che «Presso Dio il numero dei mesi [in un anno] è di dodici, iscritti nel Libro di Dio il giorno in cui creò i cieli e la terra. Di essi quattro sono sacri: tale è l'autentica legge religiosa» (traduzione italiana di Mario M. Moreno, UTET, Torino, 1967, p. 177). Il califfo 'Omar I (634-644) interpretò questo versetto come la richiesta di osservare un calendario lunare rigoroso, il quale è seguito a tutt'oggi nella maggior parte dei paesi islamici. Poiché l'anno dell'Egira è di 11 giorni circa più breve dell'anno solare, ricorrenze come il ramadan, il mese del digiuno, si spostano lentamente lungo i mesi e le stagioni, percorrendo l'intero anno nel corso di una trentina di anni solari.

Inoltre il ramadan e gli altri mesi islamici non cominciano al novilunio astronomico, definito come l'epoca in cui la Luna ha la stessa longitudine celeste del Sole ed è perciò invisibile, bensì quando si avvista per la prima volta in cielo a ponente, di sera, la sottile falce della Luna crescente. Prevedere il tempo esatto in cui la Luna crescente diventava visibile era una difficile sfida per gli astronomi matematici islamici. Benché la teoria di Tolomeo del complesso moto lunare fosse accettabilmente esatta attorno all'epoca del novilunio, essa specificava però la traiettoria della Luna solo rispetto all'eclittica (il percorso apparente del Sole sulla sfera celeste). Per prevedere la prima visibilità della Luna era necessario determinarne il moto rispetto all'orizzonte, e la soluzione di questo problema richiedeva una geometria sferica piuttosto complessa.

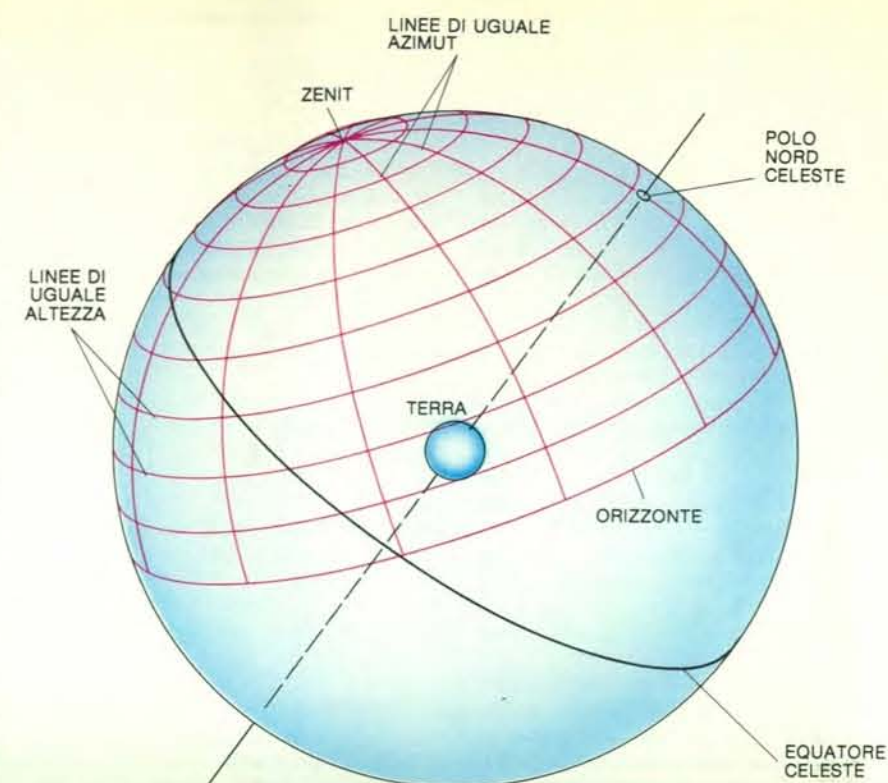
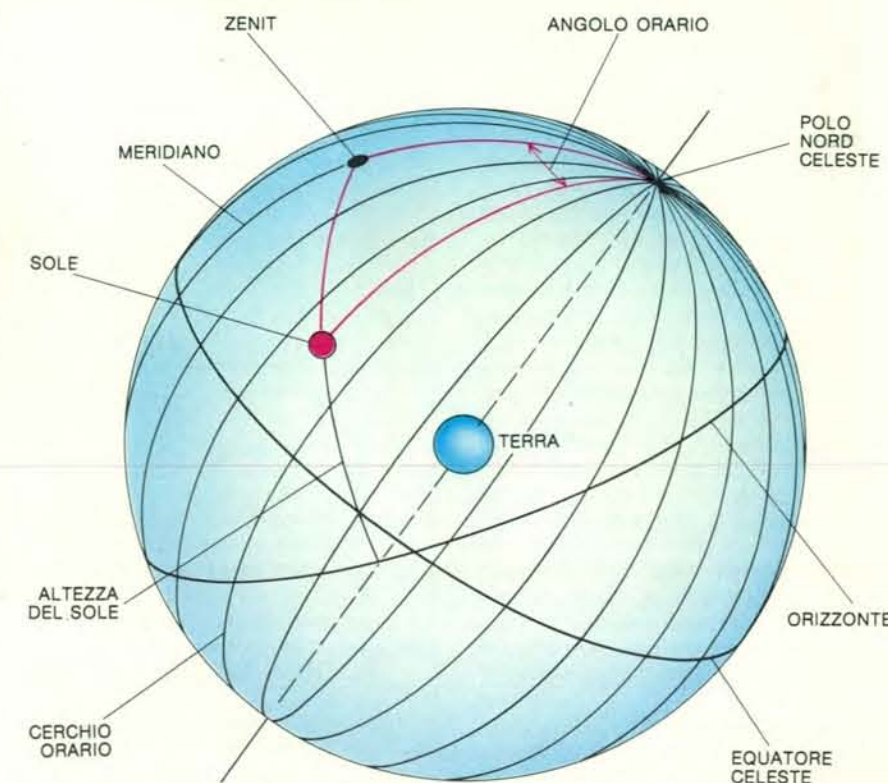
Altri due usi religiosi comportavano la soluzione di problemi che richiedevano l'applicazione della geometria sferica. Un problema, in considerazione del desiderio dei musulmani di pregare rivolti verso la Mecca e di orientare le moschee in quella direzione, era quello di determinare la direzione della città santa da qualsiasi località. Un altro problema era quello di determinare, dalla posizione osservata dei corpi celesti, l'ora giusta per le preghiere quotidiane al levar del Sole, a mezzogiorno, nel pomeriggio, al tramonto del Sole e alla sera.

La soluzione di ciascuno di questi problemi comportava che si trovassero i lati o gli angoli incogniti di un triangolo sulla volta celeste a partire da lati e angoli noti. Per esempio, un metodo per stabilire che ora è consiste nel costruire un triangolo i cui vertici siano lo zenit, il polo celeste nord e la posizione del Sole. L'osservatore deve conoscere l'altezza del Sole e quella del polo; la prima può essere osservata, e la seconda è uguale alla latitudine del luogo in cui si trova l'osservatore. L'ora è data così dall'angolo formato dall'intersezione del meridiano (il cerchio che passa per lo zenit e per il polo) con il cerchio orario del Sole (l'arco che unisce il Sole al polo).

Per risolvere i triangoli sferici Tolomeo usava un procedimento macchinoso escogitato verso la fine del I secolo da Menelao di Alessandria. Il procedimento implicava la costruzione di due triangoli rettangoli intersecantisi; applicando il teorema di Menelao era possibile trovare uno dei sei lati, ma solo se si conoscevano gli altri cinque. Per stabilire l'ora dall'altezza del Sole, per esempio, si richiedevano ripetute applicazioni del teorema di Menelao. Gli astronomi islamici erano chiaramente posti di fronte alla sfida di trovare un metodo trigonometrico più semplice.

Nel IX secolo erano state definite le sei funzioni trigonometriche moderne - seno e coseno, tangente e cotangente, secante e cosecante - mentre Tolomeo conosceva solo una funzione di una singola corda. Delle sei funzioni, pare che cinque siano state introdotte essenzialmente dagli arabi; solo la funzione del seno fu introdotta nel mondo islamico dall'India. (L'etimologia della parola «seno» è un racconto interessante. In sanscrito la parola per seno era *ardhagiya*, che significa «mezza corda», parola che in arabo fu abbreviata e translitterata come *giyb*. In arabo non si scrivono le vocali, cosicché la parola fu letta come *giayb*, che significa «tasca» o «golfo». Nell'Europa medievale essa fu quindi tradotta con *sinus*, la parola latina per «golfo» o «seno» in senso geografico.) Dopo il IX secolo lo sviluppo della trigonometria sferica fu rapido. Gli astronomi islamici scoprirono identità trigonometriche semplici, come la legge dei seni, che semplificarono e abbreviarono la soluzione di triangoli sferici.

Stelle e astrolabi

Uno fra gli esempi più vistosi dell'eredità araba entrata nell'astronomia moderna

Le coordinate altazimutali sono un modo per rappresentare l'emisfero celeste visibile a una data latitudine. Il cielo è diviso da linee di uguale altezza sopra l'orizzonte e da linee di uguale azimut attorno all'orizzonte. Queste ultime convergono allo zenit, il punto sulla verticale dell'osservatore.

La soluzione di triangoli sferici fu il nodo cruciale di molti problemi che si ponevano agli astronomi islamici. Per dedurre l'ora dall'altezza del Sole, per esempio, si dovevano trovare l'angolo orario formato al polo celeste boreale dal meridiano (il cerchio massimo che passa per lo zenit e per i poli) e il cerchio orario del Sole (il cerchio massimo che passa per il Sole e per i poli). A mezzogiorno, quando il Sole attraversa il meridiano, la sua altezza è massima e il suo angolo orario è nullo.

è rappresentato dai nomi delle stelle. Betelgeuse, Rigel, Vega, Aldebaran e Fomalhaut sono alcuni fra i nomi che hanno un'origine direttamente araba o che sono traduzioni arabe di descrizioni greche di Tolomeo.

Nell'*Almagesto* Tolomeo aveva fornito un catalogo di più di mille stelle. La prima revisione critica del catalogo fu compilata da 'Abd al-Raḥmān al-Ṣūfī, un astronomo persiano del X secolo che fu attivo sia in Iran sia a Bagdad. Il *Kitāb ṣuwar al-kawākib* («Libro delle costellazioni e delle stelle fisse») non aggiunse né sottrasse stelle rispetto all'elenco dell'*Almagesto*; al-Ṣūfī, inoltre, non misurò nuovamente le posizioni, spesso erronee, delle stelle elencate, ma fornì magnitudini più esatte, oltre a nomi arabi. Questi ultimi non erano, per la maggior parte, se non semplici traduzioni da Tolomeo.

Per molti anni si suppose che la nomenclatura delle stelle in Occidente fosse derivata dal testo arabo di al-Ṣūfī. Oggi pare invece che i suoi traduttori latini del Trecento e del Quattrocento abbiano derivato i nomi delle stelle da una versione latina dalla traduzione araba dell'*Almagesto*, combinandoli con le splendide raffigurazioni pittoriche delle costellazioni presenti nei codici di al-Ṣūfī. Frattanto la nomenclatura araba delle stelle filtrava in Occidente anche per un'altra via: la costruzione degli astrolabi.

L'astrolabio era un'invenzione greca. È essenzialmente un modello bidimensionale del cielo, un elaboratore analogico per risolvere i problemi di astronomia sferica (si veda l'articolo *L'astrolabio* di J. D. North in «Le Scienze» n. 68, aprile 1974). Un astrolabio tipico è costituito da una serie di dischi di ottone alloggiati in una matrice, anch'essa di ottone, nota in arabo come *umm* (che significa «grembo») e in italiano come «madre». Il disco superiore, chiamato in arabo *'ankabūt* («ragno») e in italiano «ragno» o «rete», è un reticolo traforato comprendente due o tre decine di punte che indicano la posizione di stelle specifiche. Inferiormente alla rete si trovano uno o più dischi piani, i climi o timpani, ciascuno dei quali reca inciso un sistema di coordinate celesti appropriato per fare osservazioni a una particolare latitudine: cerchi di uguale altezza al di sopra dell'orizzonte (analoghi ai paralleli terrestri) e cerchi di uguale azimut lungo l'orizzonte (analoghi ai meridiani terrestri). Facendo ruotare la rete attorno a un perno centrale, che rappresenta il polo celeste boreale, è possibile riprodurre il movimento diurno delle stelle sulla sfera celeste.

Benché l'astrolabio fosse noto nell'Antichità, lo strumento datato più antico che ci sia pervenuto risale al periodo islamico (*si veda l'immagine pubblicata in copertina*). Costruito da un certo Nastulus nel 315 dell'era dell'Egira (927-928 d.C.), è oggi fra gli oggetti di maggior pregio conservati nel Museo nazionale del Kuwait. Per contare gli astrolabi arabi del X secolo bastano le dita di una mano, mentre ne sono rimasti una quarantina dei secoli XI e XII. Di questi astrolabi, diversi furono costruiti in Spagna alla metà dell'XI secolo e hanno uno stile chiaramente moresco.

Il più antico trattato arabo sull'astrolabio che si sia conservato fu scritto a Bagdad da uno degli astronomi del califfo al-Ma'mūn, 'Alī ibn 'Isā. In seguito scrissero sull'astrolabio anche altri membri della scuola di Bagdad, fra i quali ha un rilievo particolare al-Farghānī. Nel trattato di quest'ultimo è sorprendente il modo matematico con il quale egli applicò lo strumento alla soluzione di numerosi problemi di astrologia, di astronomia e di computo del tempo.

Molti di questi trattati arrivarono in Spagna, dove nel XII e XIII secolo furono tradotti in latino. Il più famoso, sopravvissuto in circa 200 copie manoscritte complete, fu attribuito per molto tempo erroneamente a Māshā'allāh (Messala), un astronomo ebreo dell'VIII secolo che ebbe parte nei calcoli per la fondazione di Bagdad; probabilmente il trattato è però una stesura posteriore basata su una varietà di fonti. Attorno al 1390 questo trattato fornì la base per un saggio sull'astrolabio del poeta inglese Geoffrey Chaucer. Pare, in effetti, che l'Inghilterra sia stata la via per l'introduzione dell'astrolabio dalla Spagna nella cristianità occidentale verso la fine del Duecento e nel corso del Trecento. Non è escluso che l'attività scientifica incentrata a Oxford a quel tempo abbia contribuito al diffondersi dell'interesse per tale strumento. Il Merton College e l'Oriel College dell'Università di Oxford conservano ancora oggi pregevoli astrolabi del Trecento.

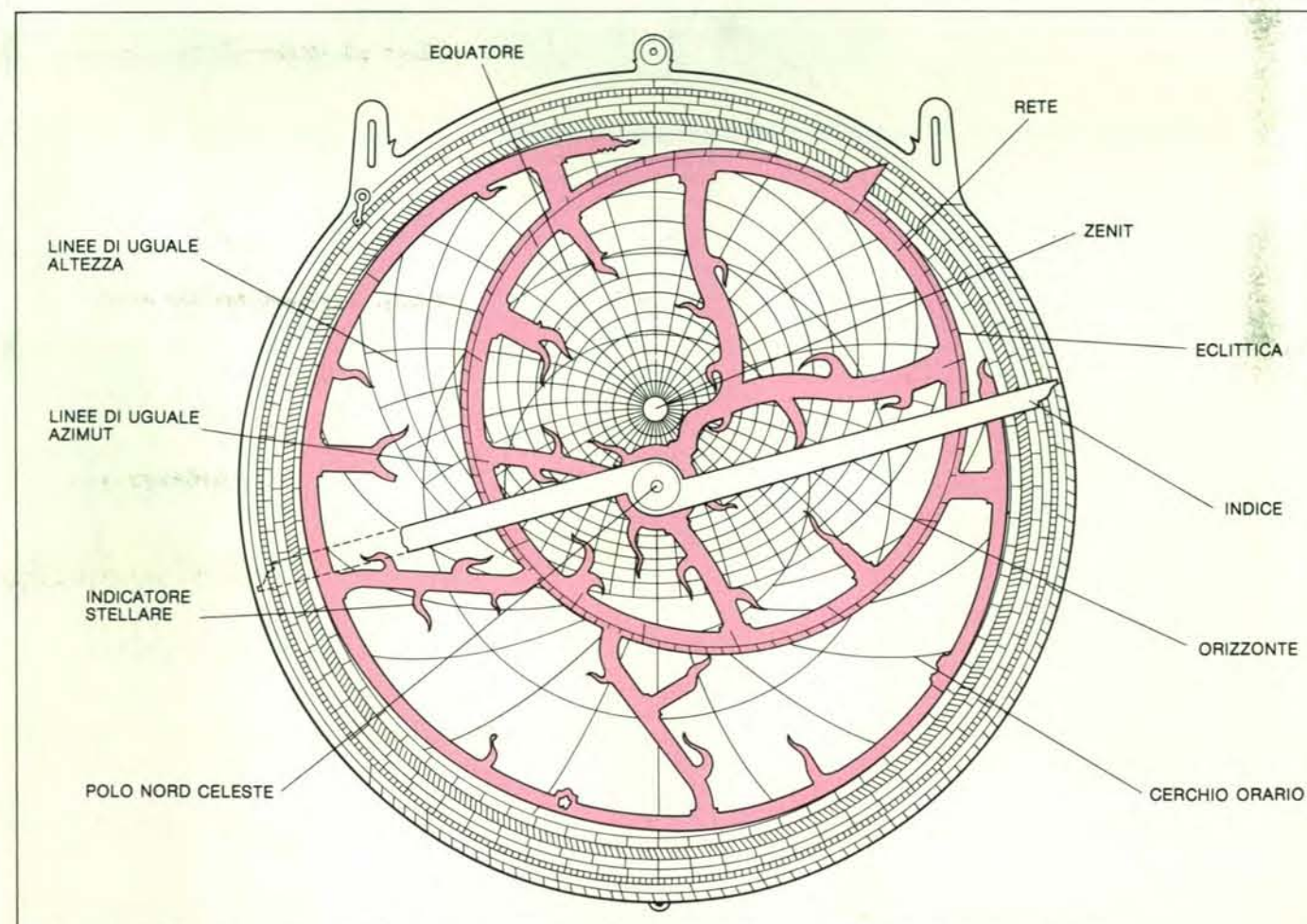
Su di essi si trovano incise tipiche serie di nomi arabi di stelle scritti in caratteri latini gotici. Sull'astrolabio del Merton College, per esempio, si leggono nomi arabi che sono entrati, sia pure con qualche modifica, nella comune nomenclatura moderna: Wega, Altahir, Algeuze, Rigil, Elfeta, Alferaz e Mirac. Così, in conseguenza della tradizione dell'astrolabio dell'Islam orientale, trasmessa attraverso la Spagna all'Inghilterra, la maggior parte delle stelle usate oggi nella navigazione ha nomi arabi, introdotti in parte dagli arabi stessi o derivati dai nomi greci di Tolomeo tradotti in arabo.

Perfezionamenti dell'opera di Tolomeo

Sarebbe sbagliato concludere, sulla base della preponderanza dei nomi arabi di stelle, che gli astronomi arabi abbiano compiuto uno studio esauriente del cielo. Al contrario, le loro osservazioni furono del tutto limitate. Per esempio, la spettacolare supernova (esplosione stellare) del 1054, che produsse la Nebulosa del Granchio, non fu registrata in testi islamici anche se trovò grande attenzione in Cina. Gli astronomi moderni, colpiti da questa vistosa omissione, spesso non si rendono conto che gli astronomi arabi trascurarono di documentare la maggior parte dei fenomeni astronomici specifici. Essi avevano poco incentivo a farlo. La loro astrologia, diversamente da quella dei cinesi, non dipendeva da presagi celesti insoliti quanto dalle posizioni dei pianeti, le quali erano descritte con sufficiente precisione dai procedimenti usati da Tolomeo.

I modelli planetari escogitati da Tolomeo nel II secolo d.C. facevano muovere il Sole, la Luna e le stelle attorno alla Terra. Una semplice orbita circolare non poteva però rendere ragione del fatto che un pianeta sembra invertire periodicamente la sua direzione di movimento in cielo. (Secondo il moderno punto di vista eliocentrico, questo moto apparente retrogrado si ha quando la Terra supera un altro pianeta, o viene superata da esso, nella sua rivoluzione attorno al Sole.) Tolomeo faceva muovere perciò ogni pianeta su un epiciclo, un cerchio rotante il cui centro si spostava attorno alla Terra lungo un cerchio maggiore detto deferente. L'epiciclo, assieme ad altri espedienti geometrici inventati da Tolomeo, forniva un'approssimazione abbastanza buona al moto apparente dei pianeti. Essendo un grande teorico, Tolomeo doveva essersi sentito abbastanza fiducioso circa la particolare geometria dei suoi modelli, giacché non si sofferma mai a spiegare in quale modo sia pervenuto alla decisione di adottarli.

D'altra parte l'idea di applicare la matematica a una descrizione numerica specifica del mondo fisico era qualche cosa di piuttosto nuovo per i greci ellenistici, e di ben lontano dalla matematica pura di Euclide e di Apollonio. In questa parte del suo programma Tolomeo deve essersi reso conto che valori migliori per i parametri numerici dei suoi modelli erano non solo desiderabili ma anche inevitabili, cosicché diede istruzioni accurate su come stabilire i parametri a partire da un numero limitato di osservazioni scelte. Gli astronomi islamici appresero questa lezione anche troppo bene. Essi limitarono le loro osservazioni, o almeno le pochissime che decisero di registrare, soprattutto a misurazioni che potessero essere usate per riderivare parametri chiave. Queste osservazioni concernevano, fra l'altro, l'orientazione e l'eccentricità dell'orbita solare e l'inclinazione del piano dell'eclittica.

L'astrolabio semplificava i calcoli astronomici, compresa la determinazione dell'ora. I suoi dischi di ottone sovrapposti sono una proiezione della sfera celeste su una superficie bidimensionale. Il disco superiore, detto rete o ragno, è un disco traforato irto di punte che indicano le posizioni delle stelle più brillanti. Il cerchio eccentrico sulla rete è l'eclittica. Sotto la rete si trova un disco pieno - un timpano o clima - su cui è inciso un sistema di coordinate celesti per una particolare latitudine; linee di uguale altezza, linee di uguale azimut e cerchi orari. Facendo ruotare la rete attorno al perno centrale, che rappresenta il polo celeste boreale, si potevano riprodurre i moti diurni delle stelle. Il primo passo per ottenere l'ora consisteva nel determinare l'altezza del Sole con l'aiuto di un'alidada e di una scala graduata sul rovescio dell'astrolabio. Poi si faceva ruotare la rete finché la posizione del Sole sull'eclittica (o dell'indicatore di una stella) si trovava in corrispondenza della giusta linea di altezza. L'angolo orario corrispondeva allora all'angolo formato al perno centrale dal meridiano e dalla linea che passava per la posizione dell'oggetto. L'astrolabio illustrato fu costruito in Inghilterra nel Duecento e appartiene al Merton College dell'Università di Oxford. Gli astrolabi furono la via principale attraverso cui i nomi arabi delle stelle passarono in Occidente. Nella fotografia sono leggibili Wega, Altahir, Algeuze, Rigil, Elfeta e Alferaz.

Un esempio particolarmente significativo di un astronomo islamico che lavorò in una cornice di riferimento rigorosamente tolemaica, stabilendo però nuovi valori per i parametri di Tolomeo, fu Muḥammad al-Bāttanī (Albategno), un contemporaneo più giovane di Thābit ibn Qurra. Le *Zīğ* («Tavole astronomiche») di al-Bāttanī sono ammirate ancor oggi come una delle opere astronomiche più importanti fra il tempo di Tolomeo e quello di Copernico. Fra l'altro, al-Bāttanī fu in grado di stabilire la posizione dell'eclittica (la traiettoria apparente percorsa in cielo dal Sole in una rivoluzione annua) con miglior precisione di Tolomeo (un compito equivalente in termini moderni a trovare la posizione dell'orbita della Terra).

Poiché al-Bāttanī non descrive le sue osservazioni nei particolari, non è chiaro se abbia adottato una strategia di osservazione diversa da quella di Tolomeo. In ogni caso ottenne buoni risultati e qualche secolo dopo i suoi parametri per l'orbita solare furono molto noti in Europa. Le sue *Zīğ* arrivarono dapprima in Spagna, dove furono tradotte in latino all'inizio del XII secolo, e in castigliano un po' più di cent'anni dopo. Il fatto che ne sia rimasta solo un'unica copia araba (nella Biblioteca dell'Escorial nei pressi di Madrid) induce a pensare che l'astronomia di al-Bāttanī non godesse nel mondo islamico di una considerazione così elevata come in Europa, dove l'avvento della stampa ne assicurò la sopravvivenza e, in particolare, la rese disponibile a Copernico e ai suoi contemporanei. Nel *De revolutionibus orbium coelestium* l'astronomo polacco cita il suo predecessore del IX secolo non meno di 23 volte.

Di contro, uno dei massimi astronomi dell'Islam medievale, 'Alī ibn 'Abd al-Raḥmān ibn Yūnus, rimase completamente ignoto agli astronomi europei del Rinascimento. Lavorando al Cairo un secolo dopo al-Bāttanī, Ibn Yūnus scrisse un importante manuale astronomico noto come *Ḥākimī Zīğ* («Tavole hakimite»). Diversamente da altri astronomi arabi, egli premise alle sue *Zīğ* una serie di più di cento osservazioni, per lo più di eclissi e di congiunzioni planetarie. Benché il manuale di Ibn Yūnus fosse molto usato nel mondo islamico, e le sue tavole per il computo del tempo siano rimaste in uso al Cairo fino al secolo scorso, la sua opera rimase totalmente sconosciuta in Occidente fino a meno di 200 anni fa.

Per l'intero periodo islamico gli astronomi se ne rimasero al sicuro all'interno della cornice di riferimento geocentrica. Per questo fatto non dovremmo criticarli troppo duramente. Fino alle osservazioni telescopiche delle fasi di Venere da parte di Galileo verso la fine del 1610 non si poté addurre alcuna prova empirica contro il sistema tolemaico. Neppure le osservazioni di Galileo riuscirono a distinguere fra il sistema geo-eliocentrico di Tycho Brahe (in cui gli altri pianeti compivano rivoluzioni attorno al Sole, mentre il Sole si muoveva attorno alla Terra) e il sistema puramente eliocentrico di Copernico (si veda l'articolo *L'affare Galileo* di Owen Gingerich in «Le Scienze» n. 170, ottobre 1982). Inoltre, benché gli astronomi islamici osservassero l'ingiunzione di Tolomeo di verificare i suoi risultati, non si limitarono semplicemente a migliorarne i parametri. I particolari tecnici dei suoi modelli non furono immuni da critiche. Questi attacchi, tuttavia, furono condotti invariabilmente sul piano filosofico e non sul piano dell'osservazione.

1

DEFERENTE
EQUANTE
EPICICLO
CENTRO
TERRA

Il sistema tolemaico poneva il Sole, la Luna e i pianeti in orbita attorno alla Terra. La Luna e i cinque pianeti noti si muovevano su cerchi sussidiari, detti epicicli, il centro dei quali percorreva cerchi maggiori chiamati deferenti. Gli epicicli riproducevano il moto retrogrado apparente dei pianeti (*a destra*). Ma la velocità orbitale di un pianeta presenta variazioni anche fuori dei cappi di moto retrogrado. Per rappresentare questo moto non uniforme, Tolomeo adottò altri due espedienti geometrici. Innanzitutto rese eccentrici i deferenti, ossia ne spostò il centro fuori della Terra, poi suppose che il moto angolare di un pianeta fosse uniforme non attorno al centro del deferente bensì attorno all'equante, un punto in posizione simmetrica alla Terra sul prolungamento della linea che unisce la Terra al centro del deferente. Alcuni astronomi islamici medievali considerarono sia gli equanti sia i deferenti eccentrici filosoficamente inaccettabili.

Dubbi su Tolomeo

I modelli di Tolomeo erano essenzialmente un sistema matematico per prevedere le posizioni dei pianeti. Eppure nelle *Hypotheseis ton planomenon* (*Ipotesi sui pianeti*) egli cercò di inquadrare i modelli in un sistema cosmologico, lo schema aristotelico di sfere strettamente racchiuse l'una nell'altra e aventi il centro in corrispondenza della Terra. Egli situò il punto più vicino della sfera di Mercurio subito dopo il punto più lontano di quella della Luna; alla massima escursione di Mercurio seguiva immediatamente il punto più vicino della sfera di Venere, e così via per le sfere del Sole, di Marte, di Giove e di Saturno.

Per riprodurre i moti apparenti non uniformi dei pianeti, Tolomeo adottò, oltre all'epiciclo, due espedienti puramente geometrici. Innanzitutto situò i cerchi deferenti in posizioni eccentriche rispetto alla Terra. In secondo luogo, fece l'assunto ingegnoso che il moto dei corpi celesti fosse uniforme non attorno alla Terra né attorno al centro dei loro deferenti, bensì attorno a un punto chiamato equante, situato sul prolungamento della linea che univa la Terra al centro del deferente, dal quale distava, in posizione simmetrica, esattamente quanto la Terra. Deferenti eccentrici ed equanti funzionavano benissimo nel rappresentare le varie velocità con le quali i pianeti sembrano muoversi in cielo, ma molti li consideravano inaccettabili dal punto di vista filosofico.

L'equante, in particolar modo, sembrava sconveniente ai filosofi, che concepivano le sfere planetarie come oggetti fisici reali, mossi ciascuno dalla sfera a esso esterna (e l'ultimo dal primo motore immobile) e che volevano essere in grado di costruire un modello meccanico del sistema. Per esempio, come sottolineò Maimonide, un dotto ebreo del XII secolo attivo in Spagna e al Cairo, il punto equante per Saturno veniva a coincidere con le sfere di Mercurio, cosa chiaramente disdicevole da un punto di vista meccanico. L'equante violava inoltre la nozione filosofica che i corpi celesti dovessero essere mossi da un sistema di cerchi perfetti, ciascuno dei quali ruotasse con velocità angolare uniforme attorno al suo centro. Per alcuni puristi persino i deferenti eccentrici di Tolomeo, che espellevano la Terra dal centro del cosmo, erano insoddisfacenti sul piano filosofico.

Gli astronomi islamici adottarono la cosmologia tolemaico-aristotelica, ma infine cominciarono a emergere alcune critiche. Una delle prime fu espressa da Ibn al-Haytham (Alhazen), attivo al Cairo nell'XI secolo e uno tra i fisici più importanti. Nei *Dubbi su Tolomeo* egli si dolse che l'equante non soddisfacesse la richiesta del moto circolare uniforme e si spinse fino a dichiarare sbagliati i modelli planetari dell'*Almagesto*.

Solo una fra le opere astronomiche di Alhazen, un libro intitolato *Sulla configurazione del mondo*, penetrò nell'Europa latina nel Medioevo. In esso Alhazen si proponeva di scoprire la realtà fisica che è alla base dei modelli matematici di Tolomeo. Concependo i cieli come sfere e gusci concentrici, egli tentò di assegnare un singolo corpo sferico a ciascuno dei moti semplici dell'*Almagesto*. L'opera fu tradotta in castigliano alla corte di Alfonso X il Saggio e poi, all'inizio del Trecento, dal castigliano in latino. Questa versione latina dell'opera, o una traduzione latina di una delle sue esposizioni, esercitò un'influenza importante in Europa all'inizio del Rinascimento. L'idea di sfere celesti separate per ogni componente dei moti planetari di Tolomeo acquistò una vasta circolazione attraverso le *Theoricae novae planetarum*, scritte dall'astronomo e matematico austriaco Georg von Peurbach attorno al 1454.

Frattanto nel XII secolo, nella regione islamica occidentale dell'Andalusia, l'astronomo e filosofo islamico Ibn Rushd (Averroè) pervenne a suo modo a una critica estrema a Tolomeo. «Affermare l'esistenza di una sfera eccentrica o di un sfera epiciclica è contrario alla natura...», scrisse. «L'astronomia del nostro tempo non offre alcuna verità, ma si accorda solo con i calcoli, non con ciò che esiste.» Averroè rifiutò perciò i deferenti eccentrici di Tolomeo e sostenne un modello rigorosamente concentrico dell'universo.

Un suo contemporaneo andaluso, Abu Isḥāq al-Biṭrūğī, si sforzò di sviluppare un tale modello rigorosamente geocentrico. I risultati furono disastrosi. Nel suo sistema, Saturno, per esempio, poteva a volte discostarsi dall'eclittica di ben 26 gradi (anziché dei tre gradi richiesti). Quanto ai moti osservati che avevano indotto Tolomeo a proporre l'equante, furono completamente ignorati. Nelle parole di un com-

Il sistema doppio-epiciclico proposto nel Trecento da Ibn-al-Shāṭir eliminò gli equanti di Tolomeo ponendo Sole, Luna e pianeti in orbite concentriche attorno alla Terra. Il disegno, da un codice conservato alla Bodleian Library dell'Università di Oxford, mostra posizioni della Luna.

mentatore moderno, al-Biṭrūgī «aggiunge caos alla confusione». All'inizio del Duecento la sua opera fu nondimeno tradotta in latino sotto il nome di Alpetragio, e a partire dal 1230 circa le sue idee furono ampiamente discusse in tutta l'Europa. Persino Copernico cita il suo ordine dei pianeti, con Venere situato oltre il Sole.

All'altro capo del mondo islamico una nuova critica dei meccanismi tolemaici fu intrapresa nel Duecento da Naṣir al-Dīn al-Ṭūsī. Fra i dotti islamici più versatili e prolifici, con 150 fra trattati e lettere noti al suo attivo, al-Ṭūsī ha anche il merito di avere costruito un importante osservatorio - che era anche un vero e proprio istituto scientifico - a Maragha (l'attuale Marāgheh, nell'Iran nordoccidentale).

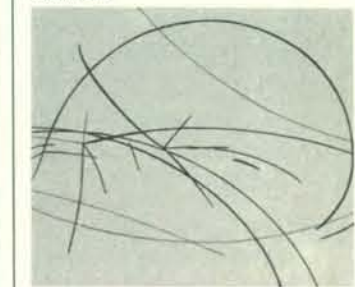
Al-Ṭūsī trovò particolarmente discutibile l'equante. In *Tadhkira* («Memorandum») lo sostituì aggiungendo altri due piccoli epicicli al modello di ciascuna orbita planetaria. Attraverso quest'espediente ingegnoso, al-Ṭūsī poté conseguire l'obiettivo di generare i moti non uniformi dei pianeti attraverso combinazioni di cerchi in rotazione uniforme. I centri dei deferenti erano però ancora esterni alla Terra. Altri due astronomi all'osservatorio di Maragha, Mu'ayyad al-Dīn al-'Urḍī e Quṭb al-Dīn al-Shīrāzī, offrirono una disposizione alternativa, ma anche questo sistema conservava l'eccentricità dei deferenti, che prestava il fianco a obiezioni di carattere filosofico.

Infine, una nuova disposizione completamente concentrica dei meccanismi planetari fu ottenuta da Ibn al-Shāṭir, che lavorò a Damasco attorno al 1350. Usando uno schema affine a quello di al-Ṭūsī, Ibn al-Shāṭir riuscì a eliminare dalle costruzioni di Tolomeo non solo l'equante, ma anche altri cerchi discutibili. Egli sgombrò perciò la via a un insieme meccanicamente accettabile di sfere celesti inserite l'una nell'altra. (Così Ibn al-Shāṭir descrisse il suo lavoro: «Trovai che i più eminenti fra gli astronomi posteriori avevano addotto dubbi incontestabili circa la ben nota astronomia delle sfere di Tolomeo. Chiesi perciò a Dio Onnipotente di ispirarmi e di aiutarmi a inventare modelli che soddisfacessero ciò che veniva richiesto, e Dio - possa Egli essere lodato ed esaltato - mi mise in grado di escogitare modelli universali per i moti planetari in longitudine e latitudine e per tutti gli altri caratteri osservabili dei loro moti, modelli sottratti ai dubbi ai quali non sfuggivano invece i modelli precedenti.») Eppure la soluzione di Ibn al-Shāṭir, come pure l'opera degli astronomi dell'osservatorio di Maragha, rimase in generale ignota in Europa.

Influenza su Copernico?

Il modello dimenticato di Ibn al-Shāṭir fu riscoperto verso la fine degli anni cinquanta di questo secolo dallo storico delle matematiche islamiche E. S. Kennedy e da suoi allievi dell'Università americana di Beirut. La scoperta sollevò un problema di grande interesse. Si riconobbe prontamente che le invenzioni di Ibn al-Shāṭir e della scuola di Maragha erano affini al meccanismo che sarebbe stato adottato qualche secolo dopo da Copernico per eliminare l'equante e per generare le complesse variazioni nella posizione dell'orbita della Terra. Copernico, ovviamente, presenta una disposizione eliocentrica dei corpi del sistema solare, ma il problema di spiegare le lente ma regolari variazioni nella velocità orbitale di un pianeta restava esattamente lo stesso. Poiché Copernico accettava le obiezioni filosofiche all'equante - alla stregua di alcuni fra i suoi predecessori islamici egli riteneva a quanto pare che i corpi celesti fossero trasportati da sfere fisiche, cristalline - tentò anch'egli di sostituire gli espedienti geometrici di Tolomeo. In un'opera preliminare, il *Commentariolus*, Copernico adottò una disposizione equivalente a quella di Ibn al-Shāṭir. In seguito, nel *De revolutionibus*, tornò all'uso delle orbite eccentriche, adottando un modello che era l'equivalente, incentrato sul Sole, di quello sviluppato a Maragha.

Copernico potrebbe aver risentito dell'influenza degli astronomi di Maragha, o di quella di Ibn al-Shāṭir? Non è stata trovata traduzione latina di nessuna delle loro opere, come pure di nessuna opera che descriva i loro modelli. Si potrebbe concepire la possibilità che Copernico abbia visto un manoscritto arabo mentre studiava in Italia (fra il 1496 e il 1503) e che se lo sia fatto tradurre, ma la cosa sembra molto improbabile. Sappiamo che una traduzione greca di parte del materiale di al-Ṭūsī giunse a Roma nel Quattrocento (molti codici greci furono trasportati in Occidente dopo la caduta di Costantinopoli avvenuta nel 1453), ma non esiste alcuna prova del fatto che essa sia mai capitata sotto gli occhi di Copernico. Oggi gli studiosi sono divisi sul problema se Copernico abbia ricevuto dal mondo islamico il suo metodo per la sostituzione dell'equante o se lo abbia escogitato in modo autonomo. È mia convinzione che possa averlo inventato indipendentemente.

Tutto lo spirito delle critiche a Tolomeo e l'idea dell'eliminazione dell'equante sono nondimeno parte del clima d'opinione che l'Occidente latino ereditò dal mondo islamico. Gli astronomi islamici sarebbero stati probabilmente sorpresi, e sarebbero forse addirittura inorriditi, se avessero potuto prevedere la rivoluzione avviata da Copernico. Eppure i motivi che indussero Copernico a introdurre il suo nuovo sistema non erano del tutto diversi dai loro. Nell'eliminare l'equante, e persino nel far rivolgere i pianeti attorno al Sole anziché attorno alla Terra, Copernico stava in parte tentando di formulare un sistema in grado di funzionare meccanicamente, un sistema che offrisse non solo una rappresentazione matematica, ma anche una spiegazione fisica dei moti planetari. In un senso profondo, Copernico non fece altro che sviluppare le implicazioni dell'astronomia fondata da Tolomeo, ma trasformata dagli astronomi islamici. Oggi tale eredità appartiene all'intero mondo della scienza.

I linfociti T e il loro recettore

Svolgono un compito chiave nel permettere all'organismo di lottare contro le infezioni virali, ma determinano anche il rigetto di un trapianto: base di questo comportamento è una molecola che è stata ormai identificata

di Philippa Marrack e John Kappler

Le cellule del sistema immunitario permettono ai vertebrati di riconoscere che materiali estranei, gli antigeni, hanno invaso il loro organismo. La risposta immunitaria che ne consegue è notevole per la sua specificità. Per esempio, un essere umano immunizzato mediante vaccinazione contro il vaiolo può opporre resistenza a un'infezione provocata da virus del vaiolo, ma non, per esempio, a un attacco di virus dell'influenza. Nell'ultimo decennio gli immunologi sono riusciti a riconoscere che il fattore più importante nella capacità del sistema immunitario di reagire specificamente ai virus è rappresentato da una classe di cellule di modeste dimensioni, i linfociti *T* (chiamati anche cellule *T*). I linfociti *T*, infatti, svolgono un ruolo ausiliare essenziale nella risposta che il sistema immunitario dà all'infezione batterica.

Da lungo tempo è stata accettata l'idea che un'attività di questo tipo venga innescata da una molecola inclusa nella membrana dei linfociti *T* e chiamata «recettore dei linfociti *T*». Si presume che un antigene specifico si adatti e si leghi a questo recettore come una chiave si adatta alla giusta serratura, mettendo quindi in moto la serie complessa di eventi biochimici che costituiscono la risposta immunitaria. Per tutta una serie di ragioni, l'isolamento del recettore dei linfociti *T* è risultato estremamente difficile e, fino a non molto tempo fa, è stato necessario dedurne le proprietà per via indiretta, senza essere guidati dalla conoscenza della sua struttura, che oggi sta per essere messa a fuoco.

Il chiarimento sulla struttura del recettore dei linfociti *T* porta con sé anche una migliore comprensione delle complesse interazioni che si instaurano tra linfociti *T* e altri elementi del sistema immunitario. In particolare, risulta sempre più evidente che il linfocito *T* è particolarmente adatto a fronteggiare le infezioni associate alle cellule dell'ospite, e non le infezioni che circolano liberamente nei liquidi corporei dell'ospite. Per svolgere questo compito, il recettore dei linfociti *T* deve non solo riconoscere un antigene specifico, ma che anche certe proteine di membrana della stessa cellula ospite. Un simile meccanismo di riconoscimento deve essere tenuto sotto rigoroso controllo perché un linfocito *T*, se dovesse essere attivato soltanto dalle proteine dell'ospite, potrebbe facilmente rivolgersi contro le cellule sane di quest'ultimo. La conseguente sensibilità del sistema di riconoscimento dei linfociti *T* è responsabile di molte affascinanti proprietà di questo sistema, importanti dal punto di vista medico. Per esempio, i linfociti *T* interverranno rapidamente nel rigetto di tessuto estraneo, chirurgicamente innestato o trapiantato nel corpo. Pertanto le ricerche sul recettore dei linfociti *T* sono di notevole interesse anche per la chirurgia.

Vi sono, a dire il vero, due tipi di linfocito responsabili del riconoscimento di antigeni specifici e il secondo tipo è costituito dai linfociti *B*. I linfociti sia *T* sia *B* sono entrambi derivati dal midollo osseo, ma i secondi subiscono un ulteriore sviluppo nel timo, una ghiandola che, nella specie umana, è situata immediatamente sotto la parte superiore dello sterno. Tutti e due i tipi circolano nel sangue e nella linfa e sono concentrati nei principali organi linfatici, che nei vertebrati superiori sono i linfonodi e la milza. Essi hanno vita lunga e, nella specie umana, possono sopravvivere anche molti anni senza dividersi. Però, come reazione a un antigene, si ingrossano molto, si dividono rapidamente e secernono numerosi fattori proteici che contribuiscono alla eliminazione dell'organismo invasore o del materiale estraneo.

È noto da tempo che la reazione iniziale di un linfocito *B* a un antigene è mediata da una proteina recettore, esposta sulla superficie della membrana cellulare. L'antigene si lega al recettore e il legame fa dividere e differenziare il linfocito *B* in un clone di plasmacellule. Queste secernono anticorpi, che presentano le medesime proprietà di legame nei riguardi dell'antigene che ha la molecola recettore inclusa nella membrana superficiale del linfocito *B* parentale. In effetti, l'anticorpo è identico al recettore del linfocito *B* al quale l'antigene si è in origine legato, tranne che in esso non si riscontra l'estremità della catena amminoacidica che àncora la proteina recettore alla membrana del linfocito *B*. Sia il recettore dei linfociti *B*, sia gli anticorpi sono chiamati immunoglobuline.

Dopo essere stati secreti nel sangue o nella linfa, gli anticorpi si legano all'antigene libero e lo contrassegnano in modo che sia poi distrutto da altri componenti del sistema immunitario. Questo quadro generale della «selezione» che un antigene fa di un linfocito *B* in base alla capacità di quest'ultimo di dare origine a un clone di cellule e di secernere anticorpi contro quell'antigene è detto «teoria della selezione clonale». La teoria è stata sviluppata negli anni sessanta da Sir Macfarlane Burnet del Walter and Eliza Hall Institute for Medical Research a Melbourne, da David W. Talmage, allora all'Università di Chicago, e da Niels Kaj Jerne, che lavorava in quel periodo per l'Organizzazione mondiale della sanità.

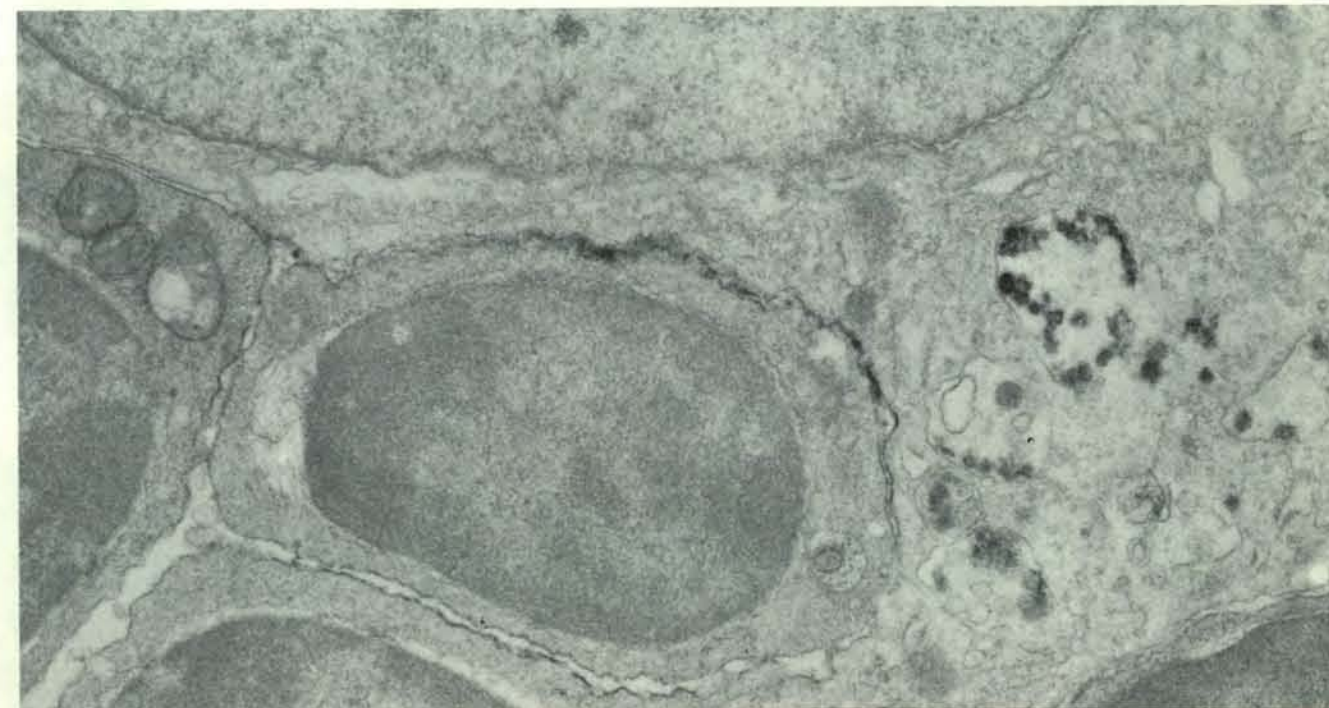
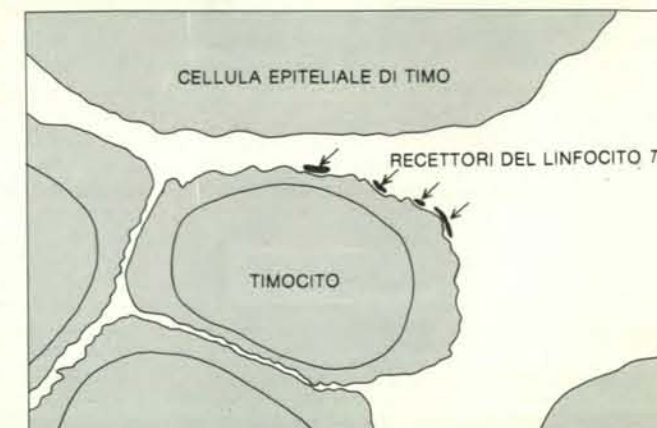
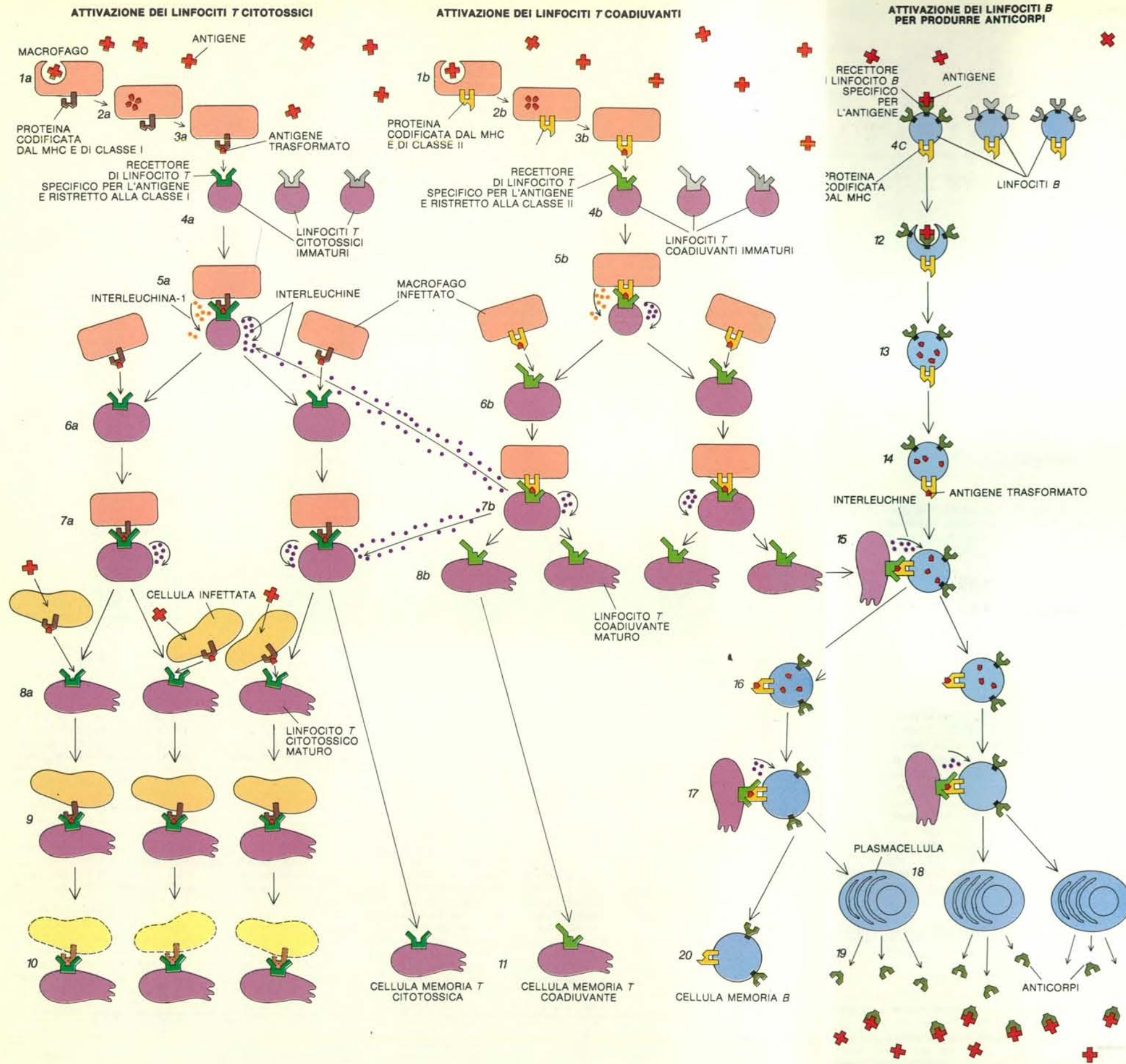
Nel ruolo di recettori dei linfociti *B*, gli anticorpi sono presenti in piccole quantità. Però, quando un linfocito *B* viene stimolato da un antigene, essi compaiono in soluzione nel siero in notevoli quantità. Sono anche secreti a livelli elevati da certi tipi di tumore dei linfociti *B* come i plasmocitomi. La pronta disponibilità di queste elevate concentrazioni di anticorpo solubile e il fatto che ogni molecola di anticorpo si possa legare a un antigene hanno permesso di isolare facilmente l'anticorpo e di acquisire notevoli conoscenze sulla struttura del recettore dei linfociti *B*.

Il recettore dei linfociti *T* è stato, invece, molto più sfuggente. Il linfocito *T*, come il linfocito *B*, reagisce a un antigene dividendosi e dando origine a un clone; poi si differenzia in uno dei numerosi tipi di linfocito *T*, specifico per quell'antigene. I linfociti *T* citotossici si legano all'antigene virale esposto sulla superficie di una cellula infetta e uccidono la cellula. I linfociti *T* soppressori inibiscono, invece, la risposta immunitaria a un antigene, un certo tempo dopo che la risposta è stata messa in atto. I linfociti *T* coadiuvanti, infine, si legano all'antigene che è presente sulla superficie di un linfocito *B*, a sua volta già legato all'antigene. Ogni linfocito *T* coadiuvante libera quindi molecole ad azione ormonosimile, le linfochine, che permettono al linfocito *B* di moltiplicarsi e di differenziarsi. Così, per liberare l'enorme potenziale distruttivo di un linfocito *B* esiste un sistema a due chiavi: una (l'antigene libero) per il recettore dei linfociti *B* e l'altra (l'antigene sulla superficie del linfocito *B*) per il recettore dei linfociti *T*.

I linfociti *T* non si differenziano mai in cellule che secernono anticorpi. Pertanto, diversamente dal recettore dei linfociti *B*, il loro recettore non è facilmente disponibile nelle quantità di composto chimico solubile e puro che sono necessarie per svolgere un'analisi soddisfacente. Essendo gli anticorpi «costruiti» in maniera così raffinata ed essendo essi così efficienti nel riconoscimento degli antigeni, per molti anni si è supposto che i linfociti *T* si basassero sulle stesse molecole dei linfociti *B* per legarsi e per reagire all'antigene.

Molti ricercatori hanno speso anni per esaminare le superfici dei linfociti *T* e i loro secreti alla ricerca di immunoglobuline. Benché ricerche ad ampio raggio abbiano suggerito che i linfociti *T* soppressori possono legarsi all'antigene con molecole simili alle immunoglobuline, molti esperimenti hanno provato che, per la maggior parte, i linfociti *T* non sono associati a immunoglobuline. Non solo le immunoglobuline non vengono secrete da questi linfociti, ma esse non si trovano neppure sulle loro membrane o nel loro citoplasma. Altri esperimenti dimostrano che i linfociti *T* non esprimono l'RNA messaggero trascritto dai geni per le immunoglobuline; inoltre quando, nei linfociti *T*, si esaminano i geni per le immunoglobuline si trova che essi non subiscono riassestamenti come capita nei linfociti *B*.

Questi risultati negativi si sono accumulati negli anni settanta e agli inizi degli anni ottanta. I risultati di esperimenti come quelli citati non sono però, di solito, molto persuasivi quando sono presi in considerazione singolarmente. Tuttavia, il numero complessivo degli insuccessi ha fatto sì che gli esperti di immunologia molecolare cambiassero gradatamente il modo di impostare la ricerca. Se non era possibile trovare delle immunoglobuline associate ai linfociti *T*, occorreva cercare altrove, in queste cellule, il recettore per l'antigene.

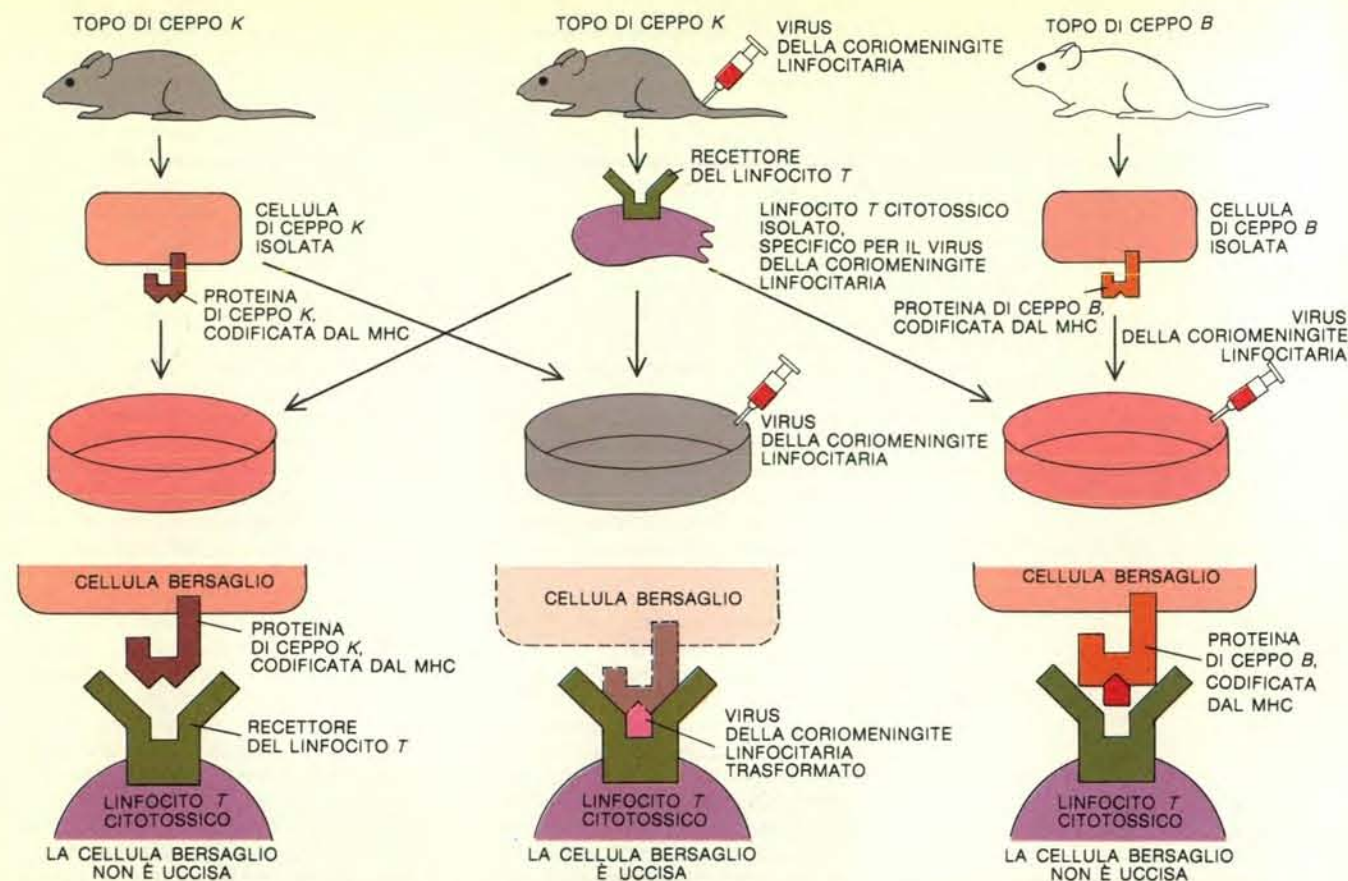
Alcuni anni prima, lo studio del trapianto di tessuti estranei e del rigetto da parte dell'animale ospite aveva portato alle stesse conclusioni gli esperti di immunologia cellulare: per esempio, le cellule prelevate da un ceppo di topi vanno incontro a un rapido rigetto quando sono iniettate in un topo appartenente a un diverso ceppo genetico. A partire dagli inizi degli anni trenta, Peter Gorer e altri ricercatori hanno dimostrato che un simile rigetto è provocato da molecole di antigene, presenti sulla superficie delle cellule di topo estranee.

La «istruzione» di un linfocito *T* nel timo gli conferisce la capacità di distinguere le altre cellule dello stesso organismo da cellule estranee. Questa capacità è un elemento d'importanza critica nella specificità della risposta immunitaria. Il timocito, o linfocito *T* immaturo, che si nota nella microfotografia, appare ingrandito 20 900 diametri ed è unito a una cellula epiteliale che fa parte del rivestimento del timo (*si veda lo schema a destra*). I siti di legame scuri, presenti sulla superficie del timocito, sono le molecole dei recettori: esse probabilmente si legano a proteine che definiscono l'«io» e che sono esposte sulla cellula epiteliale del timo. Queste proteine sono codificate nel genoma da un segmento di DNA chiamato maggior complesso di istocompatibilità (MHC). Il legame può stimolare l'inizio della «istruzione», o maturazione, del timocito. I linfociti *T* maturi acquisiscono una scarsa affinità per le proteine del proprio io, codificate dal MHC, mentre hanno una affinità elevata per queste stesse proteine quando sono associate ad antigeni. I linfociti *T* innescano una risposta immunitaria solo contro le cellule infettate del proprio io, oppure contro cellule appartenenti a un individuo estraneo. La microfotografia è stata ottenuta da Andrew G. Farr dell'Università di Washington.

Le proteine che caratterizzano ogni cellula come «propria» o come «estranea» sono codificate da geni associati e localizzati in una regione del DNA chiamata il maggior complesso di istocompatibilità (MHC dall'inglese *major histocompatibility complex*). Esse sono note come «proteine codificate dal MHC», o semplicemente «proteine MHC». Una delle loro proprietà più sorprendenti è l'estremo polimorfismo: vi sono milioni di alleli, o varianti, dei geni del MHC, che codificano per ogni proteina; pertanto, la probabilità che due individui non affini abbiano identiche proteine MHC è davvero scarsa.

Grazie alla ricerca di Gorer, è risultato sempre più chiaro che sia il rigetto dei tessuti trapiantati sia l'esistenza di proteine codificate dal MHC sono in stretta relazione con la risposta immunitaria. Sir Peter Medawar e altri ricercatori hanno dimostrato che responsabili del riconoscimento di molecole antigeniche su un tessuto estraneo trapiantato sono i linfociti; ricerche successive hanno indicato che fondamentali nel rigetto di tessuti sono i linfociti *T*, e non i linfociti *B*. Tuttavia, i trapianti non sono esperimenti della natura, e gli immunologi avevano quindi ancora dubbi sul ruolo delle proteine codificate dal MHC.

Alcuni indizi sulle funzioni svolte dalle proteine MHC cominciarono a emergere alla metà degli anni sessanta. Hugh O. McDevitt, allora al National Institute for Medical Research, in Inghilterra, assieme a Michael Sela del Weizmann Institute of Science, in Israele, e Baruj Benacerraf, allora al National Institute of Allergy and Infectious Diseases, con i suoi collaboratori hanno studiato la risposta di vari ceppi di animali agli antigeni e hanno trovato che quella di un topo a certi antigeni sintetici dipende dal ceppo genetico a cui quel topo appartiene. Per esempio, quando il polimero TGAL viene iniettato nei topi, quelli che possiedono alleli *b* nel loro MHC producono anticorpi anti-TGAL, mentre quelli che possiedono alleli *k* no. È risultato così subito evidente che gli alleli in questione agiscono sulle funzioni dei linfociti *T* e non su qualche altro tipo cellulare del sistema immunitario.

A questo punto alcuni immunologi hanno suggerito l'idea che il recettore dei linfociti *T* potesse essere codificato dal MHC. Se così fosse davvero, ci si aspetterebbe che differenti tipi di MHC codifichino per differenti recettori dei linfociti *T* e quest'ultimo punto spiegherebbe le differenze rilevate da un ceppo di topi all'altro nella risposta al TGAL. La teoria ha però dovuto essere scartata quando si è scoperto che alcune delle proteine codificate dal MHC, che influiscono sul legame del TGAL, non sono neppure espresse alla superficie dei linfociti *T* di topo. Pertanto, la fonte delle differenze osservate non poteva essere solo il recettore dei linfociti *T* ed era inutile continuare a supporre che esso fosse codificato dal MHC. Invece, gli esperimenti hanno fornito una conferma a una questione più sottile: le differenze rilevate nella risposta immunitaria al TGAL (produzione di anticorpi specifici) indicavano che le proteine codificate dal MHC influiscono sul modo in cui quel polimero viene riconosciuto dal recettore del linfocito *T*. Questo risultato è stato il primo indizio che il presupposto per l'attività del recettore fosse un'interazione tra antigeni estranei e proteine codificate dal MHC.

Come poter spiegare questa strana collezione di risultati? Le risposte sono state date dalle ricerche pionieristiche di Ethan M. Shevach e Alan S. Rosenthal del

Esistono tre tipi principali di risposta del sistema immunitario, specifica nei riguardi dell'antigene: la risposta di un linfocito *T* citotossico (*a sinistra*), quella di un linfocito *T* coadiuvante (*al centro*) e quella di un linfocito *B* (*a destra*). L'antigene che penetra nel corpo (*in alto*) viene fagocitato da un macrofago (*1a*, *1b*), trasformato in frammenti proteici più corti (*2a*, *2b*) ed esposto in forma parzialmente digerita alla sua superficie (*3a*, *3b*). In questa sede esso si lega a una delle due classi di proteine codificate dal MHC. Un linfocito *T*, il cui recettore si adatta al complesso proteina-antigene, si lega a quest'ultimo e, pertanto, è «selezionato» per la propagazione clonale dal complesso proteina-antigene (*4a*, *4b*). Analogamente, l'antigene libero seleziona un linfocito *B*, il cui recettore si adatta a esso e i due si legano (*4c*). I linfociti *T* destinati a diventare citotossici si legano all'antigene associato con proteine codificate dal MHC e di classe I (*5a*), mentre i futuri linfociti *T* coadiuvanti si legano all'antigene associato con proteine di classe II (*5b*). Il legame con il linfocito *T* stimola la liberazione da parte del macrofago della interleuchina-1, una molecola ormonosimile; il macrofago, a sua volta, stimola il linfocito *T* a dividersi e a differenziarsi (*6a*, *6b*). La divisione cellulare prosegue fino a quando il linfocito *T* è stimolato da cellule che espongono antigeni (*7a*, *7b*). Un linfocito *T* maturo (*8a*, *8b*) può allora svolgere la sua funzione. Se è citotossico, può o legarsi a una cellula infetta che espone l'antigene (*9*) e ucciderla (*10*), oppure rimanere in circolo nel sangue o nella linfa come cellula memoria, la quale agisce più prontamente con qualsiasi antigene dello stesso tipo che, in futuro, entrerà in contatto con essa (*11*). Anche un linfocito *T* coadiuvante maturo può diventare una cellula memoria. La funzione del linfocito *T* coadiuvante consiste nello stimolare la proliferazione di linfociti *B* attivati. Un linfocito *B*, che ha inglobato il suo antigene legato (*12*) e lo ha trasformato (*13*), presenta anche in superficie un frammento di antigene legato a una proteina codificata dal MHC e di classe II (*14*). Il linfocito *T* coadiuvante maturo può legarsi a questo punto con il complesso antigene-proteina, presente sul linfocito *B* (*15*). Questo legame libera le interleuchine, che permettono al linfocito *B* di dividersi e di differenziarsi (*16*); la divisione cellulare prosegue fintantoché il linfocito *B* viene stimolato dai linfociti *T* coadiuvanti (*17*). Le plasmacellule mature (*18*) liberano quindi i loro recettori, specifici per gli antigeni, come anticorpi e questi si legano agli antigeni liberi distruggendoli (*19*). Altri linfociti *T* maturi, presenti nel clone, rimangono in circolo come cellule memoria (*20*). Lo schema include i risultati degli esperimenti descritti dagli autori in questo articolo assieme ai risultati di molti altri esperimenti, ma certi particolari del modello sono ancora oggetto di vivaci controversie.

Un linfocito *T* reagisce a un antigene solo se legato a una proteina codificata dal MHC, esposta su una cellula bersaglio. Linfociti *T* di topo di ceppo *k*, immunizzato contro il virus della coriomeningite linfocitaria, possono uccidere le cellule che espongono proteine di ceppo *k*, codificate dal MHC, dopo che sono state infettate dal virus (*al centro*). Non uccidono, invece, le cellule infettate dal virus che non espongono le proteine (*a destra*) e neppure le cellule di topo che espongono le proteine di ceppo *k*, codificate dal MHC, ma non sono infettate dal virus (*a sinistra*).

National Institute of Allergy and Infectious Diseases, di David H. Katz e Benacerraf della Harvard Medical School, di Bernice Kindred dell'Università di Costanza nella Repubblica Federale Tedesca, di Donald Shreffler della School of Medicine dell'Università del Michigan e, in modo estremamente chiaro, dagli importanti esperimenti di Rolf Zinkernagel e Peter C. Doherty della Australian National University.

Nel corso di un esperimento, Zinkernagel e Doherty hanno iniettato dosi subletali del virus della CML (coriomeningite linfocitaria) in topi e hanno isolato linfociti *T* citotossici, specifici per quel virus (*si veda l'illustrazione di questa pagina*), dagli animali immunizzati. La funzione abituale di questi linfociti consiste nel riconoscere gli antigeni virali sulle cellule infettate e nel distruggere queste ultime. La maggior parte dei ceppi di topo produce anticorpi contro il virus della CML e, pertanto, vi erano scarse probabilità che quel virus non riuscisse a indurre in alcuni ceppi di topo una risposta immunitaria simile a quella scatenata dal TGAL. Malgrado questo fatto, Zinkernagel e Doherty hanno trovato con loro sorpresa che i linfociti *T* di topi appartenenti al ceppo *k* e immunizzati contro la CML sono in grado di uccidere le cellule infettate dal virus della CML solo se esse contengono almeno uno dei numerosi tipi di proteine codificate dal MHC e caratteristiche del ceppo *k*. In altre parole, i linfociti *T*, letali per le cellule infettate che contenevano queste proteine, non potevano uccidere cellule di topi affini, infettati con lo stesso virus, ma dotati, per esempio, di proteine codificate dal MHC e caratteristiche del ceppo *d*.

Questo esperimento e altri ancora hanno dimostrato che i linfociti *T* citotossici erano interessati non solo agli antigeni virali delle cellule infettate, ma anche alle sequenze amminoacidiche delle proteine codificate dal MHC e presenti su quelle cellule. Altri ricercatori hanno dimostrato poco dopo che la stessa cosa avveniva con i linfociti *T* citotossici, che avevano una specificità per altri antigeni. Il fenomeno è oggi definito «restrizione da MHC». In generale, le proteine codificate dal MHC e riconosciute dai linfociti *T* citotossici appartengono a una singola classe strutturale, le proteine della classe I, che si trovano sulla superficie di tutte le cellule nucleate dell'organismo.

Noi e diversi altri ricercatori abbiamo quindi trovato che anche i linfociti *T* coadiuvanti presentano questa restrizione da MHC; in generale, però, un linfocito *T* coadiuvante riconosce una differente classe di proteine codificate dal MHC, le proteine di classe II. L'espressione di queste proteine è notevolmente meno diffusa nell'organismo dell'espressione delle proteine della classe I: negli esseri umani, le proteine della classe II si trovano soltanto nei linfociti *T* e *B*, nei macrofagi e in certe cellule appartenenti ad altri tessuti.

Non appena è stata scoperta la restrizione riguardo all'MHC, è risultato chiaro che i fatti potevano essere spiegati da due tipi di teoria. La teoria del riconoscimento associato suggerisce nella sua forma estrema che ogni linfocito *T* porta recettori di un unico tipo. Ciascuno di questi recettori si legherebbe in qualche modo a un complesso tra l'antigene e una particolare proteina codificata dal MHC, che compare sulla superficie della cellula che espone quell'antigene. Per contro, la teoria del duplice riconoscimento suggerisce nella sua forma estrema che ogni linfocito *T* presenta due tipi di recettore, uno dei quali si legherebbe a un antigene specifico e l'altro a una proteina specifica, codificata dal MHC.

Ogni tipo di teoria presenta vantaggi e svantaggi. La teoria del riconoscimento associato spiega l'osservazione in base alla quale i linfociti *T* si legano solo raramente o all'antigene o alle proteine codificate dal MHC senza legarsi contemporaneamente a entrambe le molecole. La teoria spiega anche bene i risultati ottenuti da McDevitt, Sela e Benacerraf sulle proprietà immunogene del TGAL nei topi. Se un singolo recettore di linfocito *T* deve simultaneamente legare l'antigene e la proteina codificata dal MHC, l'incapacità dei topi di ceppo *k* di reagire al TGAL potrebbe essere dovuta al fatto che l'antigene TGAL non riesce a formare un complesso con la proteina codificata dal MHC e di ceppo *k*. Dal canto suo, la teoria del duplice riconoscimento non presenta uno svantaggio che si riscontra nella teoria del riconoscimento associato: cioè essa non deve postulare un'interazione tra ogni forma di antigene e le forme invarianti della proteina codificata dal MHC che si trovano in ogni singolo animale.

Benché l'ultima parola non sia stata ancora detta, è probabilmente giusto affermare che i dati oggi disponibili sono in favore della teoria del riconoscimento associato, con il suo modello di un unico recettore del linfocito *T* per l'antigene e per la proteina codificata dal MHC. Per esempio, abbiamo studiato le proprietà di linfociti *T* ibridi che esprimono recettori derivati da due differenti linfociti *T* parentali (*si veda l'illustrazione a pagina 34*). Abbiamo dapprima isolato dei linfociti *T* coadiuvanti, che riconoscono in maniera specifica un antigene, l'ovalbumina di pollo, quando è associato a certe proteine di classe II, tipiche del ceppo *k*, e li abbiamo fusi con un tumore costituito da linfociti *T*, così da ottenere un ibridoma di linfociti *T*, cioè una cellula ibrida moltiplicatasi rapidamente e che, in coltura, ha dato origine in poco tempo a cloni.

Per ottenere una cellula con due differenti insiemi di recettori di linfociti *T*, abbiamo fuso una delle cellule del nostro ibridoma di linfociti *T* con un secondo insieme di linfociti *T*, specifici per un certo antigene. Questo secondo insieme poteva reagire soltanto all'antigene costituito dall'emocianina del mollusco *Fissurella* in presenza di proteine di classe II e di ceppo *f*.

Abbiamo analizzato la risposta di linfociti *T* appena fusi agli antigeni ovalbumina di pollo ed emocianina di *Fissurella*, associati a cellule che espongono gli antigeni e che appartengono o al ceppo *k* o al ceppo *f*. Se la teoria del duplice riconoscimento fosse giusta, i linfociti *T* appena fusi dovrebbero presumibilmente portare quattro tipi di recettore, uno per ogni antigene e uno per ogni proteina codificata dal MHC e appartenente alla classe II. Pertanto essi dovrebbero reagire a entrambi gli antigeni esposti su un linfocito di ceppo *k* e a entrambi gli antigeni esposti su un linfocito di ceppo *f*. Abbiamo trovato, invece, che essi reagiscono agli antigeni solo come hanno fatto in precedenza i loro linfociti parentali: all'antigene ovalbumina di pollo, associato con linfociti di ceppo *k*, e all'antigene emoglobina di *Fissurella*, associato con linfociti di ceppo *f*. Non vi era alcun cenno di reazione all'ovalbumina di pollo nei linfociti di ceppo *f* o all'emoglobina di *Fissurella* nei linfociti di ceppo *k*.

Se la teoria del riconoscimento associato è giusta, l'antigene e la proteina codificata dal MHC devono in qualche modo formare un complesso prima che un linfocito *T* possa legarsi a essi. I risultati di diversi esperimenti suggeriscono che tra le proteine codificate dal MHC e gli antigeni esista un qualche tipo di interazione; tuttavia solo pochi esperimenti sono riusciti a rivelarla direttamente. Forse la dimostrazione più sorprendente negli ultimi tempi è quella offerta dalle ricerche di Emil R. Unanue e collaboratori della School of Medicine della Washington University.

Unanue e collaboratori hanno identificato un piccolo antigene che può essere riconosciuto dai linfociti *T* quando è associato con proteine codificate dal MHC e appartenenti al ceppo *k*, ma non quando è associato a proteine del ceppo *d*. Essi hanno posto uguali concentrazioni dell'antigene in soluzione su entrambi i lati di una membrana semipermeabile e hanno aggiunto da un solo lato proteine codificate dal MHC e appartenenti o all'uno o all'altro ceppo. Dato che le molecole antigeniche sono molto più piccole di entrambi i tipi di proteina, la membrana permetteva il libero passaggio degli antigeni, ma confinava le proteine nel loro compartimento iniziale.

Unanue e il suo gruppo hanno trovato che, quando le proteine codificate dal MHC e appartenenti al ceppo *k* venivano aggiunte a un compartimento e la concentrazione degli antigeni veniva poi lasciata equilibrare, gli antigeni apparivano in maggiore quantità nel compartimento che conteneva le proteine del ceppo *k*. Le concentrazioni dell'antigene nei due compartimenti non si modificavano quando, a un compartimento, venivano aggiunte proteine codificate dal MHC e appartenenti al ceppo *d*. Il risultato suggerisce che l'antigene si leghi, almeno qualche volta, alla proteina codificata dal MHC assieme alla quale è riconosciuto, anche se il processo può avvenire con un'efficienza non elevata.

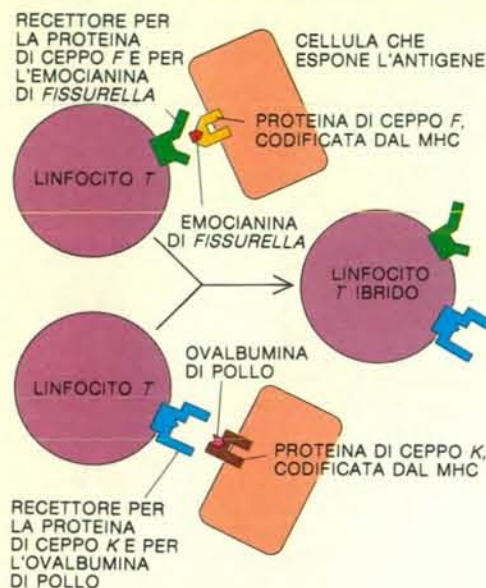
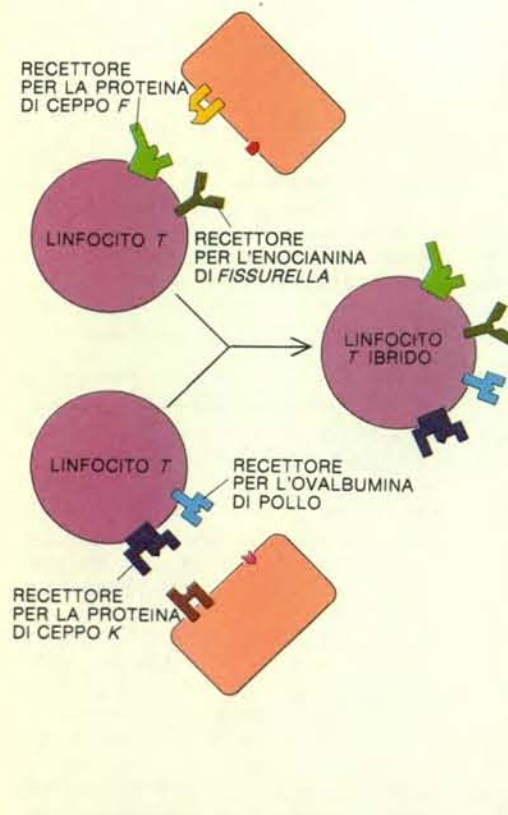
Anche dopo che erano state stabilite tutte le proprietà del recettore dei linfociti *T*, citate sopra, l'identità della proteina responsabile dell'attività del recettore rimaneva ancora avvolta dal mistero. Tuttavia, fu chiaro da quel momento che il recettore dei linfociti *T* differisce per almeno un aspetto importante dalla molecola dell'anticorpo. Indipendentemente dal fatto che esso sia un'unica proteina distinta o due, perlomeno qualche sua componente ha una forte tendenza a riconoscere parte di una proteina codificata dal MHC. La molecola dell'anticorpo no.

Vi sono stati due importanti progressi tecnologici che hanno permesso di cominciare a intravvedere la struttura del recettore dei linfociti *T*. Uno è stata la scoperta effettuata agli inizi degli anni settanta da George Köhler e Cesar Milstein del Laboratory of Molecular Biology del Medical Research Council a Cambridge, i quali hanno trovato un modo per produrre ibridomi di linfociti *B*, che possono secernere *in vitro* grandi quantità di tipi di anticorpo selezionati. Questi anticorpi sono definiti monoclonali e possono essere prodotti e purificati in enormi quantità. Il secondo progresso è consistito nella elaborazione di metodi grazie ai quali cloni di linfociti *T* o ibridomi di linfociti *T* con specificità per determinati antigeni e proteine codificate dal MHC potevano essere propagati in coltura. Noi e numerosi altri ricercatori abbiamo sfruttato entrambe queste conquiste per produrre anticorpi contro il recettore dei linfociti *T* e quindi per identificare le proprietà strutturali di questo recettore.

Per produrre anticorpi contro il recettore dei linfociti *T* abbiamo innanzitutto «costruito» un ibridoma di linfociti *T* che portavano un recettore per il riconoscimento dell'ovalbumina di pollo, antigene associato a una proteina di classe II, appartenente al ceppo *d*. I recettori presenti in questo tipo di ibridomi, quando sono legati con il giusto antigene e la giusta proteina codificata dal MHC, danno una rapida risposta, secernendo linfochine. Gli ibridomi sono utili dal punto di vista sperimentale perché la produzione di linfochine può essere indotta e misurata rapidamente in coltura.

Per ottenere un anticorpo di specificità nota si deve immunizzare un animale contro un antigene noto. Nel caso del recettore dei linfociti *T* la precisa identità dell'antigene non era, tuttavia, conosciuta. La strategia della nostra ricerca è consistita nell'immunizzare topi contro ibridomi di linfociti *T*, nella speranza di indurre la produzione di anticorpi che interferissero solo con le funzioni presumibilmente svolte dai recettori su quegli ibridomi. Abbiamo pensato che un anticorpo contro il recettore si sarebbe legato a quest'ultimo bloccandone la capacità di fissarsi all'ovalbumina di pollo e alla proteina di ceppo *d*, codificata dal MHC. Una diminuzione nella produzione di linfochine indicherebbe il blocco.

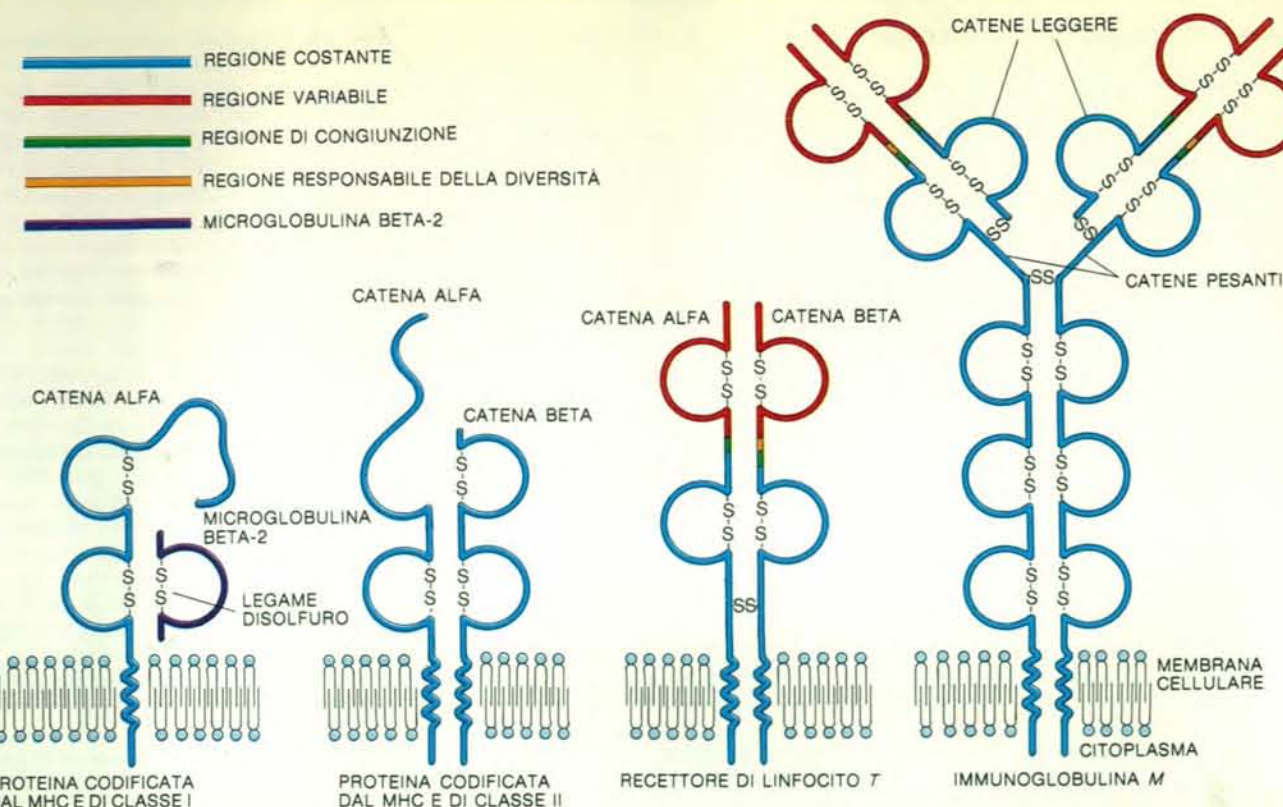
Abbiamo immunizzato molti topi contro un numero elevato di ibridomi di linfociti *T*; quindi abbiamo prelevato il siero da ogni topo, in tempi diversi a partire dall'immunizzazione, e lo abbiamo analizzato in coltura allo scopo di valutare la sua capacità di bloccare la reazione degli ibridomi di linfociti *T* alle proteine di ceppo *d* e all'ovalbumina di pollo. Alla fine abbiamo identificato molti topi che producevano l'anticorpo bloccante. L'antisiero prelevato da questi animali aveva un'ulteriore proprietà incoraggiante: esso non bloccava la risposta dei linfociti *T* quando ibridomi di questi linfociti con differenti specificità venivano posti, in coltura, a confronto con l'antigene e con la proteina codificata dal MHC, nei riguardi dei quali erano specifici. Abbiamo ripetuto l'esperimento con altri ibridomi di linfociti *T* nel ruolo dell'ibridoma originario. In ogni caso gli anticorpi bloccanti prodotti dai topi erano efficaci solo contro l'ibridoma di linfociti *T* utilizzato per l'immunizzazione.

La specificità degli antisieri ottenuti per un particolare ibridoma di linfociti *T* ci ha fatto ritenere che eravamo davvero riusciti a produrre anticorpi contro un particolare

Due teorie tra loro incompatibili possono però spiegare il riconoscimento sia dell'antigene sia delle proteine codificate dal MHC da parte dei linfociti *T*. Secondo la teoria del riconoscimento associato, ogni linfocito *T* porta un recettore, che può legarsi a un complesso di un antigene con una proteina codificata dal MHC (*in alto a sinistra*). Secondo la teoria del duplice riconoscimento, invece, ogni linfocito *T* ha due recettori, uno per l'antigene e uno per la proteina codificata dal MHC (*in basso a sinistra*). Per comprovare le due teorie, gli autori hanno «costruito» un linfocito *T* ibrido fondendo due linfociti *T* normali. Un linfocito *T* parentale è stato isolato da animali con proteine di ceppo *f*, codificate dal MHC, dopo che erano stati immunizzati contro l'emocianina di *Fissurella*. Un secondo linfocito *T* parentale è stato isolato, invece, da animali di ceppo *k*, immunizzati contro l'ovalbumina di pollo. Secondo una versione estrema della prima teoria, il linfocito ibrido dovrebbe riconoscere l'emocianina di *Fissurella* solo su cellule di ceppo *f* e l'ovalbumina di pollo solo su cellule di ceppo *k* (*in alto a destra*). Secondo una versione estrema della seconda teoria, l'ibrido dovrebbe riconoscere, invece, i due antigeni sulle cellule di entrambi i ceppi (*in basso a destra*). Il linfocito ibrido si è comportato proprio conformemente a quanto sostiene la prima teoria.

recettore dei linfociti *T*. Questo recettore è l'unica struttura sulla superficie dell'ibridoma di linfociti *T*, che presumibilmente sarebbe bloccato da un anticorpo in un modo così specifico, perché soltanto esso varierebbe da un clone di linfociti *T* a un altro. Ciononondimeno il nostro compito non era ancora finito. Dato che solo piccole quantità di anticorpi potevano venire isolate da ogni topo, per «immortalare» le cellule che secernevano questi anticorpi abbiamo fatto ricorso al metodo di Köhler e Milstein.

Kathryn Haskins e Janice White del nostro laboratorio hanno estratto plasmacellule *B* da uno dei topi immunizzati contro un ibridoma di linfociti *T*, specifici per l'antigene ovalbumina di pollo. Le plasmacellule sono state poi fuse con cellule tumorali in modo da ottenere ibribomi di plasmacellule che sarebbero stati facilmente coltivati *in vitro*. Questi ibridomi hanno secreto anticorpi, i quali sono stati selezionati in base alla capacità di bloccare il riconoscimento dell'ovalbumina di pollo e della proteina di classe II e di ceppo *d* da parte dell'ibridoma di linfociti *T*, specifici per l'ovalbumina di pollo. Un solo ibridoma di plasmacellule ha mostrato questa proprietà. Analoghi anticorpi sono stati ottenuti all'incirca nello stesso periodo da James P. Allison, allora al Cancer Center dell'Università del Texas a Smithville, negli Stati Uniti, e da Stefan C. W. Meuer e Ellis Reinherz della Harvard Medical School.

Grazie soprattutto alla riuscita propagazione degli anticorpi per i recettori specifici dei linfociti *T* è stato possibile costruire un'immagine di questi recettori proteici. Il solido legame tra recettore e anticorpo ha fatto sì che i ricercatori potessero purificare il recettore in grandi quantità, sufficienti a mettere in luce le sue proprietà molecolari di base. Cosa che forse può sorprendere, esiste a quanto pare una stretta rassomiglianza strutturale tra la molecola dell'anticorpo e il recettore dei linfociti *T*. Entrambi sono costituiti da due catene polipeptidiche codificate da geni ben distinti nel DNA e tenuti uniti da forti legami covalenti che collegano due atomi di zolfo (*si veda l'illustrazione della pagina a fronte*). Negli anticorpi le due catene differiscono per la dimensione e sono chiamate catena pesante e catena leggera; ciascuna include una sequenza di amminoacidi relativamente costante in un singolo animale, perfino per anticorpi che si legano ad antigeni distinti. Ogni catena ha però anche un lungo tratto di amminoacidi che varia in misura considerevole negli anticorpi con diversa specificità antigenica.

La variabilità di un anticorpo all'interno di un animale deriva dal genoma di quest'ultimo. L'idea di fondo è che ogni catena polipeptidica dell'anticorpo consta di tre o quattro regioni, ciascuna delle quali può essere codificata da uno o dall'altro di diversi segmenti di DNA, selezionati a caso. La variabilità combinatoria che insorge nella «costruzione» dell'anticorpo porta

Le strutture molecolari della proteina di classe I e di quella di classe II, codificate dal MHC, e inoltre del recettore del linfocito *T* e di una immunoglobulina o anticorpo, sono simili; esse hanno anche in comune alcune sequenze di amminoacidi e sono caratterizzate da anse costituite, all'interno di ogni catena, da circa 70 amminoacidi; gli atomi di zolfo, a ogni estremità delle anse, sono uniti da legami covalenti. Le proteine di classe I, nei vertebrati superiori, sono espresse alla superficie di ogni cellula nucleata in associazione con la beta-2-microglobulina, una proteina non codificata dal MHC. Le proteine di classe II sono espresse, invece, solo sulla superficie di cellule selezionate, come i linfociti *B*. Nel recettore del linfocito *T* e nella immunoglobulina sono state identificate una regione variabile, una di congiunzione e una responsabile della diversità, ognuna codificata da una sequenza di DNA, selezionata a caso in mezzo a parecchie sequenze presenti nel genoma. Queste regioni danno origine a una sostanziale variabilità combinatoriale nelle molecole espresse da differenti cellule. Gli schemi, molto semplificati, non sono disegnati in scala.

a un numero elevato di anticorpi ben distinti tra loro.

Per esempio, una catena pesante consta di quattro regioni, cioè di quattro sequenze amminoacidiche: la regione costante, una regione di congiunzione, una regione responsabile della diversità e la regione variabile. La regione costante è codificata da un solo segmento di DNA, mentre il segmento di DNA che codifica per la regione di congiunzione può essere selezionato a caso tra quattro segmenti distinti. Analogamente la regione responsabile della diversità è codificata da un segmento di DNA selezionato tra più di dieci e la regione variabile è codificata da un segmento di DNA selezionato tra più di 100. Il numero di sequenze amminoacidiche che possono aver origine dalle varie combinazioni è, pertanto, superiore a $4 \times 10 \times 100$, o 4000. Anche la catena leggera consta di regioni codificate da differenti combinazioni di DNA e una ulteriore variabilità trae origine dal fatto che i punti di contatto tra la regione di congiunzione, quella responsabile della diversità e la regione variabile sono imprecisi. Secondo Susumu Tonegawa del Massachusetts Institute of Technology (MIT), il numero di molecole di anticorpo diverse può raggiungere il miliardo (si veda l'articolo *Le molecole del sistema immunitario* di Susumu Tonegawa in «Le Scienze» n. 208, dicembre 1985).

Nel recettore dei linfociti *T*, le due catene sono chiamate catena alfa e catena beta. Nel topo esse hanno entrambe un peso molecolare di circa 43 000 unità di massa atomica (dalton); nella specie umana la catena alfa pesa circa 50 000 unità di massa atomica e la catena beta circa 39 000. Confrontando le catene alfa e le catene beta estratte da differenti cloni di linfociti *T*, abbiamo trovato che su ogni catena, di clone in clone, compaiono certe sequenze fisse di amminoacidi. Altre sequenze, invece, possono variare.

Con la scoperta delle proteine che costituiscono il recettore dei linfociti *T*, sono entrate in gioco le tecniche della biologia molecolare. Queste tecniche permettono ai ricercatori di analizzare i geni che codificano per una data proteina e un'analisi di questo tipo può chiarire la struttura di una proteina molto più rapidamente di quanto possano fare i metodi biochimici. Inoltre, dati i ruoli analoghi dei recettori dei linfociti *T* e delle immunoglobuline nel sistema immunitario e le somiglianze strutturali che stanno emergendo tra loro, è sembrato probabile che i segmenti di DNA che codificano per il recettore dei linfociti *T* subissero un riassestamento prima che le proteine che costituiscono il recettore si esprimessero sulla superficie del linfocito *T*, proprio come accade per i segmenti di DNA che codificano per il recettore dei linfociti *B*. Si è scatenata quindi subito una gara per trovare i geni responsabili del recettore dei linfociti *T*.

I primi attendibili candidati al ruolo di geni sono stati comunicati simultaneamente da due gruppi: Stephen M. Hedrick, Mark M. Davis e collaboratori dei National Institutes of Health, dell'Università della California a San Diego e della Stanford University, e Tak W. Mak e collaboratori dell'Ontario Cancer Institute. Entrambi i gruppi, il primo lavorando su geni di topo e il secondo su geni umani, pensavano che le proteine del recettore dei linfociti *T* sarebbero state trovate soltanto nei linfociti *T* e non, per esempio, nei linfociti *B*. Per confermare questa loro supposizione, essi hanno adottato tecniche sperimentali raffinate e sono stati presto in grado di identificare i geni espressi solo nei linfociti *B* o solo nei linfociti *T*, ma non in entrambi. Sono stati identificati dapprima i geni che codificano per la catena beta del recettore e, nell'anno successivo, sono stati isolati anche quelli per la catena alfa. Dall'analisi di questi geni si sono potuti

conoscere numerosi particolari sulla struttura del recettore dei linfociti *T*.

Sia la catena alfa, sia la catena beta del recettore di questi linfociti hanno la regione variabile, la regione costante e la regione di congiunzione. Inoltre, i ricercatori sopra citati hanno confermato la presenza di una regione responsabile della diversità nella catena beta e probabilmente di una regione analoga anche nella catena alfa. Le sequenze amminoacidiche di ciascuna di queste regioni sono simili alle sequenze analoghe presenti nelle immunoglobuline, ma non sono affatto identiche.

Davis e collaboratori da un lato e Leroy Hood e collaboratori del California Institute of Technology dall'altro hanno preso in esame l'organizzazione delle sequenze di DNA che codificano per la catena beta del recettore e hanno trovato finora 12 segmenti di DNA che codificano per altrettante regioni di congiunzione, due segmenti che codificano per regioni responsabili della diversità (ognuno di questi segmenti può essere trascritto secondo l'uno o l'altro di tre sistemi di lettura) e circa 20 segmenti che codificano per regioni variabili. Il numero delle possibili combinazioni di amminoacidi nella catena beta è, pertanto, almeno 12 × 2 × 3 × 20, cioè 1440. Ma anche questo numero è troppo piccolo: una stima realistica della variabilità deve anche tenere conto delle mutazioni e di una congiunzione inesatta, che possono influire sul recettore dei linfociti *T*, proprio come influiscono sulla molecola di anticorpo. La variabilità nella catena alfa può anche essere superiore: benché non siano state identificate finora regioni responsabili della diversità, sembrano esservi più regioni di congiunzione e regioni variabili di quante ve ne siano per la catena beta; forse anche 100 per ogni tipo. Le catene alfa e beta possono dunque combinarsi per formare differenti tipi di recettore dei linfociti *T*, nell'ordine di 10 milioni, il che è sufficiente a giustificare il ben noto repertorio di linfociti *T* presente in un animale.

Le nuove informazioni acquisite sulla struttura del recettore dei linfociti *T* hanno permesso di riproporre in termini ben più precisi molte delle questioni riguardanti il sistema immunitario. Vi sono essenzialmente tre tipi di riconoscimento che sono stati osservati e per i quali si vorrebbe ora dare una spiegazione a livello strutturale. In primo luogo, il linfocito *T* non reagisce alle proteine del proprio «io», codificate dal MHC. In secondo luogo, il linfocito *T* reagisce quando si trova simultaneamente in presenza di un antigene e di una proteina del proprio io, codificata dal MHC, ma generalmente non quando si trova in presenza di un antigene associato a una proteina codificata dal MHC, ma appartenente a un altro individuo. In terzo luogo, il linfocito *T* può reagire a una proteina di un altro individuo, codificata dal MHC, se in assenza dell'antigene. È questo l'effetto che spiega il rigetto di tessuti innestati o trapiantati. Associata a queste osservazioni vi è una questione inerente lo sviluppo: in che modo i precursori dei linfociti *T* si differenziano nel timo in cellule che hanno queste proprietà?

La spiegazione più lineare della tolleranza immunologica è che i cloni di linfociti *T* che reagiscono alle proteine del proprio io, codificate dal MHC, sono eliminati nel timo in un modo o nell'altro. Non vi è ancora una descrizione precisa di come una simile selezione potrebbe aver luogo. Si potrebbe suggerire che, a un certo stadio dello sviluppo, i linfociti *T* muoiono se i loro recettori, che hanno una specificità per un determinato antigene, si legano a una qualsiasi molecola espressa dalle cellule dell'organismo a cui essi appartengono. Questa spiegazione è nota come «teoria dell'aborto clonale». Essa sostiene che, allorché i linfociti *T* si sviluppano, tutti quelli che reagiscono in maniera specifica alle proteine del proprio io, codificate dal MHC, oppure ad altri antigeni del proprio io, verrebbero uccisi perché continuamente bombardati da quelle molecole. Morirebbero anche i linfociti *T* che reagiscono in modo specifico alle proteine del proprio io, codificate dal MHC e associate a un antigene estraneo, se dovessero legarsi a simili complessi, mentre in un animale non colpito da un antigene di questo tipo essi si svilupperebbero fino a raggiungere la piena maturità. Potrebbero così sopravvivere fino a quando un antigene estraneo che invadesse l'organismo scatenasse la loro reazione.

La risposta del linfocito *T* a una proteina codificata dal MHC e associata a un antigene fa sorgere delle perplessità. Rolf Zinkernagel e collaboratori e Michael J. Bevan, allora al MIT, hanno riconosciuto innanzitutto le dimensioni del problema nel corso di uno studio da essi intrapreso sulla «istruzione» dei linfociti *T* nel timo (*si veda l'illustrazione qui sotto*). Essi hanno incrociato topi appartenenti ai ceppi *a* e *b*, generando così una progenie ibrida che conteneva nel proprio organismo proteine codificate dal MHC e appartenenti a entrambi i ceppi. Il timo di ogni topo ibrido è stato quindi asportato e l'animale è stato successivamente irradiato allo scopo di distruggere tutti i suoi linfociti *B* e *T*. Un nuovo timo è stato quindi trapiantato in esso da un topo di ceppo *b* e nuove cellule staminali del midollo osseo di un altro animale ibrido dello stesso tipo sono state parimenti trapiantate.

In un topo ibrido normale le cellule staminali ibride si svilupperebbero in linfociti *T* maturi in grado di reagire collettivamente all'antigene, associato o con una proteina codificata dal MHC e appartenente al ceppo *a*, o con una proteina codificata dal MHC e appartenente al ceppo *b*. Negli animali da laboratorio, tuttavia, sulle cellule non linfatiche del timo erano presenti soltanto proteine di ceppo *b*, anche se, nel resto del corpo, si trovavano, invece, proteine caratteristiche di entrambi i ceppi. È stato trovato con sorpresa che i linfociti *T* maturi di questi animali potevano reagire all'antigene soltanto se questo era associato a una proteina appartenente al ceppo *b* e codificata dal MHC, e non se era associato, invece, a una proteina di ceppo *a*. Così, allorché i linfociti *T* si sviluppavano nel timo di animali appartenenti al ceppo *b*, i loro recettori venivano apparentemente selezionati per riconoscere soltanto quegli antigeni che erano associati alle proteine codificate dal MHC, presenti in quell'organo. A quanto pare i linfociti *T* non riescono a riconoscere gli antigeni associati a proteine codificate dal MHC ed estranee al timo nel quale sono ospitati.

Sono state avanzate molte ipotesi per spiegare questo inatteso risultato, ma nessuna è del tutto soddisfacente. Forse, perché i linfociti *T* possano maturare e diventare funzionanti, i loro recettori devono legarsi debolmente alle proteine del proprio io, codificate dal MHC e presenti nel timo. Successivamente quei linfociti *T*, i cui recettori hanno la massima affinità per le proteine del proprio io e codificate dal MHC, potrebbero andare incontro a un aborto clonale; i linfociti *T* che rimangono, con una affinità scarsa ma ancora positiva per le proteine del proprio io e codificate dal MHC, verrebbero allora riversati in circolo. Quando, però, un antigene si lega a una di queste proteine, l'affinità del recettore presente sui linfociti *T* circolanti per quell'antigene potrebbe aumentare in maniera sostanziale.

Che dire dell'intensa reazione dei linfociti *T* a un tessuto estraneo trapiantato? La spiegazione che trova maggior favore è che a un linfocito *T* una proteina estranea, codificata dal MHC, sembra chimicamente la stessa cosa del complesso tra una proteina del proprio io, codificata dal MHC, e un antigene. Questa spiegazione tiene conto di parecchie osservazioni. Per esempio, il recettore del linfocito *T* che si lega alla proteina del proprio io, codificata dal MHC, sembra essere lo stesso recettore che anche si lega alla proteina estranea. Inoltre, questo recettore sembra avere una predisposizione a legarsi all'antigene in presenza o della proteina di classe I o della proteina di classe II, ma non di entrambe, indipendentemente dal fatto che la proteina codificata dal MHC sia o no estranea.

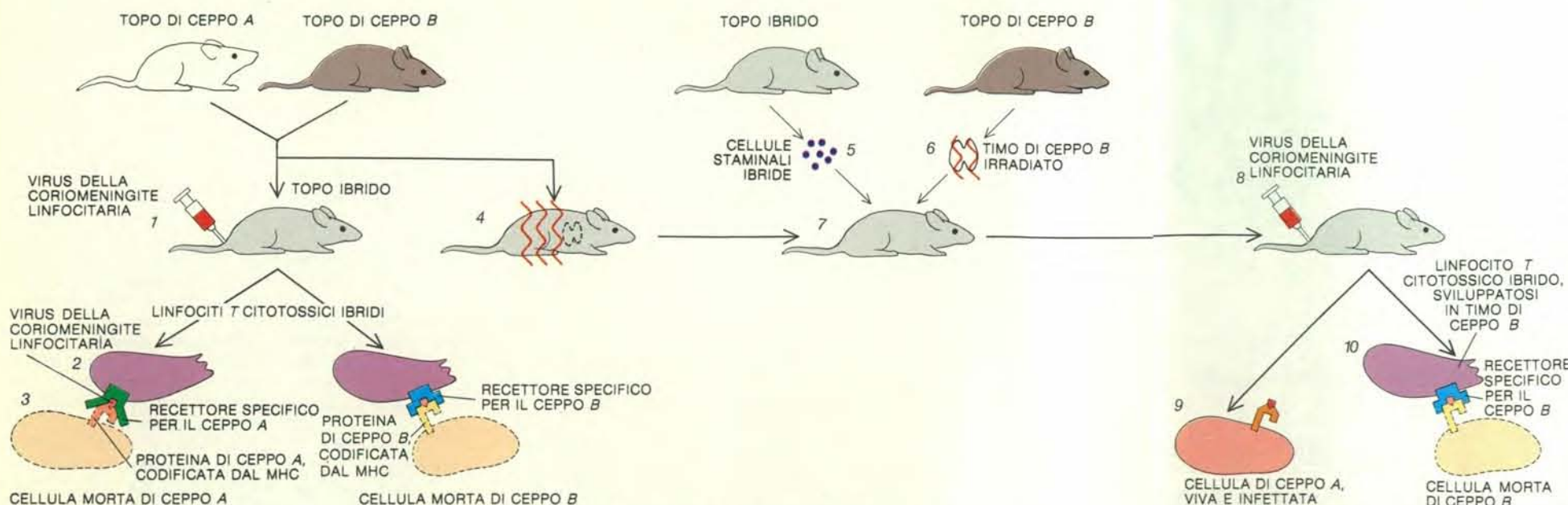
Nell'ultimo anno si sono svolte numerose ricerche per trovare un fondamento a queste predisposizioni del recettore. Finora, però, non sono state osservate differenze ovvie tra i linfociti *T* ristretti all'una o all'altra classe di proteine codificate dal MHC. Nel 1984, Tonegawa e collaboratori hanno trovato un terzo tipo di gene, il gene gamma, che subisce una riassestamento da parte dei linfociti *T*. Le proprietà della sequenza di questo gene fanno pensare che la proteina che il gene codifica non faccia parte del recettore del linfocito *T*, così come questo recettore viene correntemente inteso. Cionondimeno la proteina si esprime solo nei linfociti *T* citotossici e così contribuisce, forse, in qualche modo non sospettato alla specificità, nei riguardi del MHC, della cellula che la contiene.

Se esiste una somiglianza sul piano chimico tra una proteina codificata da un MHC estraneo e un complesso costituito da una proteina codificata dal MHC del proprio io e un antigene, si potrebbe pensare che l'antigene che fa parte del complesso sia piccolo. Rispetto a un antigene grosso, un antigene piccolo si adatterebbe anche più facilmente alla fenditura, nel recettore del linfocito *T*, che funge da legame. Recenti ricerche condotte da numerosi gruppi, tra cui quelli di Emil Unanue e collaboratori e di Ronald H. Schwartz e collaboratori del National Institute of Allergy and Infectious Diseases e il nostro, in collaborazione con Howard M. Grey del National Jewish Center for Immunology and Respiratory Medicine, hanno confermato che l'antigene legato è piccolo.

Per esempio, abbiamo trovato che i linfociti *T* reagiscono a cellule vive, tenute in incubazione assieme a un antigene, mentre non rispondono allo stesso antigene quando esso viene aggiunto alle cellule dopo che sono state fissate con una sostanza chimica. Quando alle cellule fissate con una sostanza chimica si aggiunge solo un piccolo frammento dell'antigene, il linfocito *T* reagisce di nuovo. La ricerca conferma che le cellule che espongono l'antigene, e alle quali può legarsi un linfocito *T*, hanno già trasformato, o digerito, in qualche modo l'antigene.

Dobbiamo sollevare un'ultima obiezione. L'esistenza dei linfociti *B* dimostra chiaramente che si può sviluppare un sistema in cui un recettore riconosca un antigene nativo e libero e si leghi a esso. Questo legame finisce per portare a un'espansione e a un differenziamento clonali dei linfociti *B*, il che dà origine ad anticorpi efficaci nel distruggere l'antigene. Perché, allora, i linfociti *T* hanno messo in atto un sistema così elaborato per riconoscere l'antigene solo in associazione con le proteine prodotte dal MHC?

Si può dare a questa domanda una risposta teleologica: il linfocito *T* è programmato solo per reagire con un antigene fissato su una cellula bersaglio, e non con un antigene libero. Per esempio, un linfocito *T* citotossico è programmato per uccidere cellule infettate da un virus e, pertanto, inibisce la crescita e la diffusione del virus. Un linfocito del genere non può uccidere una particella virale libera e, pertanto, le sue risorse potrebbero venir sperperate se non vi fosse il mezzo per dirigere la sua attenzione su un virus che si sviluppa all'interno di una cellula nucleata. Così il sistema immunitario ha programmato il recettore del linfocito *T* in modo che possa legarsi a un antigene virale solo quando può legarsi anche a una proteina codificata dal MHC del proprio io e appartenente alla classe I, proteina che è presente sulla superficie di ogni cellula nucleata.

Il sistema immunitario, ancor più astutamente, ha progettato dei recettori sui linfociti *T* coadiuvanti, i quali sono destinati a interagire, in primo luogo, con i linfociti *B* e con altre cellule del sistema immunitario. Questi recettori si legano all'antigene solo quando esso è associato con proteine codificate dal MHC del proprio io e appartenenti alla classe I, proteine che vengono espresse solo alla superficie dei linfociti *B* e delle altre cellule immunitarie. Il sistema immunitario si è così evoluto in modo che entrambe le classi delle proteine codificate dal MHC fungano da segnale per i linfociti *T*, guidando questi ultimi verso l'antigene esattamente nei punti dove la loro azione possa essere efficace.

L'effetto del timo sul linfocito *T* è stato dimostrato in un esperimento di Rolf Zinkernagel, allora alla Scripps Clinic and Research Foundation. Topi ibridi figli di genitori di ceppo *a* e di ceppo *b* sono serviti come controlli. Essi sono stati immunizzati contro il virus della coriomeningite linfocitaria (*1*) e hanno poi fornito linfociti *T* citotossici, specifici per il virus (*2*). Questi linfociti sono stati in grado di uccidere le cellule infettate dal virus e provenienti da topi sia di ceppo *a* sia di ceppo *b* (*3*). Per determinare il ruolo del timo, quest'organo è stato asportato da un altro topo ibrido e tutti i linfociti *T* preesistenti, assieme ad altre cellule linfoidi, sono stati distrutti irradiando l'animale (*4*). Nuove cellule staminali del midollo osseo, da cui normalmente si sviluppano i linfociti *T*, sono state allora ottenute da un altro topo ibrido (*5*) e il timo irradiato di un topo di ceppo *b* è stato trapiantato nell'ibrido timectomizzato (*6*). Le cellule staminali ibride sono state fatte sviluppare nel topo ibrido con timo di ceppo *b* (*7*). Quando questo topo è stato immunizzato contro il virus della coriomeningite linfocitaria (*8*), i linfociti *T* isolati da esso non sono stati in grado di uccidere le cellule infettate dal virus e derivate da topi di ceppo *a* (*9*), anche se erano in grado di uccidere le stesse cellule appartenenti ai topi di ceppo *b* (*10*).

Sono prevedibili i terremoti?

Una risposta certa sarà possibile solo quando il meccanismo fisico che è all'origine dei fenomeni sismici sarà così ben compreso da consentire la determinazione di alcuni parametri fondamentali dell'evento da prevedere

di Enzo Boschi e Michele Dragoni

Come l'eclisse di Sole, il terremoto è un evento improvviso, non preceduto normalmente da segnali visibili a tutti e in base ai quali la gente possa identificare l'imminenza dell'evento. Mentre però l'eclisse è prevedibile con grande precisione da secoli, il terremoto non lo è. Da che cosa dipende questa differenza? E sarà possibile in futuro prevedere i terremoti? Per rispondere a queste domande, è necessario anzitutto chiarire che cosa si intende per «previsione». Questo termine viene infatti usato in più di un senso dai sismologi, a seconda del contesto al quale si riferisce. Si parla di previsione statistica e di previsione deterministica e ancora di previsione a lungo, medio e breve termine. Ne segue che i discorsi sulla previsione si prestano talvolta a equivoci, specialmente quando sono rivolti ai non addetti ai lavori.

Per previsione di un terremoto intendiamo qui la determinazione di un certo numero di parametri che caratterizzano l'evento sismico: tali parametri devono comprendere almeno la posizione dell'epicentro, l'istante di tempo in cui ha inizio il terremoto (tempo origine) e la sua magnitudo (o una grandezza equivalente). Poiché la precisione delle nostre misurazioni è sempre limitata, i valori di questi parametri avranno necessariamente una certa indeterminazione. La previsione sarà tanto migliore quanto più l'indeterminazione sarà piccola. Per terremoti medio-grandi (magnitudo maggiore di 6), potremmo considerare accettabili indeterminazioni di qualche decina di chilometri per l'epicentro, di alcuni giorni per il tempo origine, di mezza unità per la magnitudo (che corrisponde a un fattore cinque circa nell'energia liberata sotto forma di onde sismiche). Diremo pertanto che una previsione è verificata se si produce un evento con caratteristiche (epicentro, tempo origine, magnitudo) i cui valori sono all'interno dei rispettivi intervalli specificati in anticipo. Dal punto di vista pratico, una previsione, oltre che sufficientemente precisa, deve essere formulata con sufficiente anticipo da poter mettere in atto efficaci misure di prevenzione dei danni.

Per fare comprendere le difficoltà che si incontrano nelle ricerche miranti alla previsione dei terremoti, è necessario accennare brevemente al meccanismo che produce gli eventi sismici. Il guscio esterno della Terra, la litosfera, è frammentato in una dozzina di grandi zolle (o placche), più molte altre minori. A causa dei moti che avvengono all'interno della Terra, ancora non del tutto compresi, queste zolle si muovono lentamente l'una rispetto all'altra, con velocità relative di alcuni centimetri all'anno. Naturalmente il moto di ciascuna zolla è ostacolato dalle zolle adiacenti, per quello che, in termini semplici, possiamo chiamare un effetto di «attrito» lungo i margini delle zolle stesse. Ciò fa sì che certi segmenti dei margini, bloccati dall'attrito per tempi più o meno lunghi, vengano deformati dal moto delle zolle, cosicché vi si accumula un grande sforzo elastico. A un certo punto, tale sforzo supera la resistenza di attrito e il segmento di margine ha un brusco scorrimento, che gli fa recuperare in pochi secondi lo spazio di cui era rimasto indietro. È questo scorrimento che genera nella Terra le onde elastiche che percepiamo come terremoto.

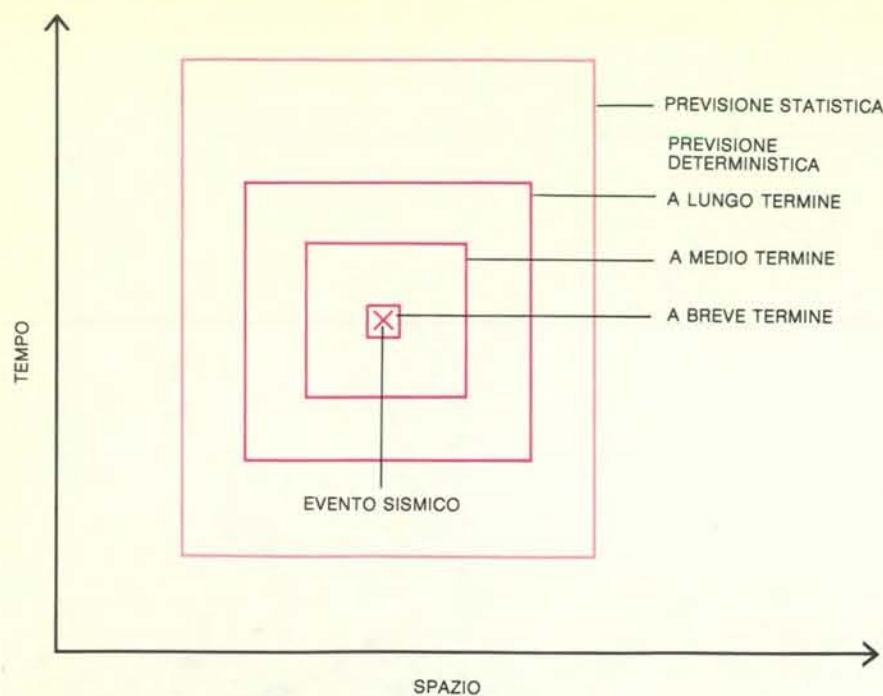
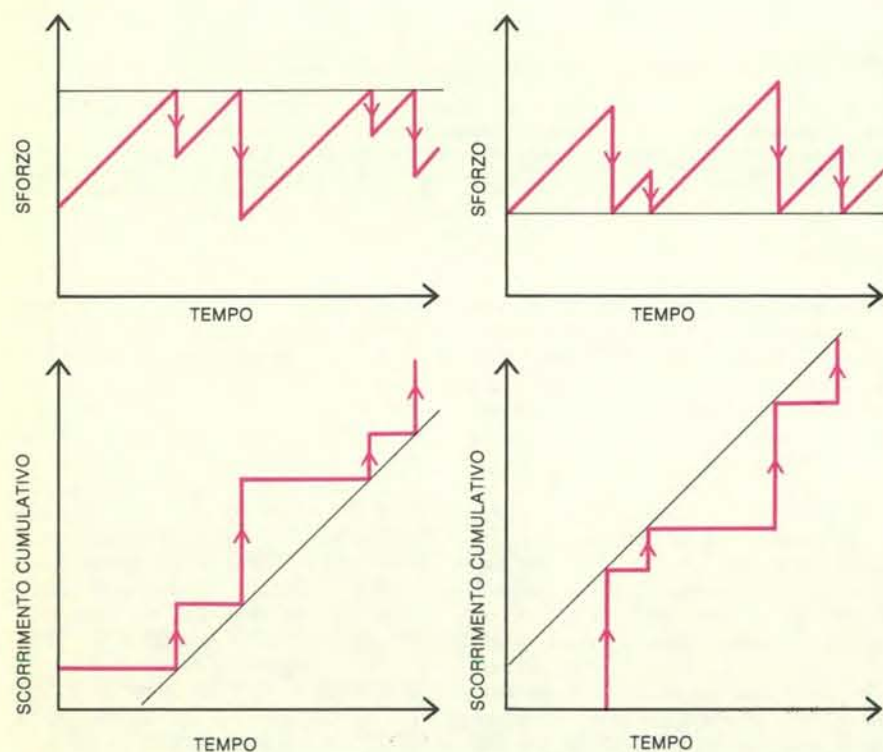
Nella realtà la situazione è assai più complessa di quella descritta. Tra l'altro, i margini delle zolle non sono in generale lineamenti ben definiti alla superficie terrestre, ma appaiono come fasce intensamente deformate e fratturate, larghe decine o centinaia di chilometri. All'interno di queste fasce, gli scorrimenti che producono i terremoti avvengono su superfici di minore resistenza, dette faglie. Le faglie sono dunque le sorgenti dei terremoti. Lo scorrimento ha inizio in un punto della faglia e si propaga rapidamente su un'area che, nei terremoti più grandi, è di migliaia di chilometri quadrati.

Detto questo, sottolineiamo che prevedere i terremoti con gli strumenti della scienza oggi non è possibile. Non esistono «schemi operativi» per la previsione. La previsione dei terremoti è però oggetto di ricerca. Si può dire che qualunque ricerca, mirante alla comprensione dei processi che sono all'origine dei terremoti, ci fa compiere passi avanti nella direzione della prevedibilità. Le osservazioni geofisiche hanno consentito di individuare tutta una serie di fenomeni «precursori» dei terremoti. Tra questi, i principali sono una lenta deformazione del suolo, che può protrarsi per anni prima del terremoto; variazioni nell'attività sismica di fondo normalmente presente nelle regioni sismiche; variazioni nella velocità di propagazione delle onde sismiche nell'area in seguito colpita dal terremoto. E ancora: variazioni di resistività elettrica delle rocce, anomalie del campo magnetico terrestre, emissioni di gas rado dalle rocce, variazioni del livello e della temperatura dell'acqua nei pozzi profondi. In alcuni casi, sono state registrate onde sismiche a frequenza molto bassa (10^{-3} hertz), che sono state attribuite a processi analoghi a quelli che producono i terremoti, ma molto più lenti, chiamati appunto «terremoti lenti» o «silenziosi». Alcuni fenomeni precursori sono stati riprodotti in laboratorio con esperimenti su campioni di roccia. Quando una roccia viene sottoposta a sforzo, essa si deforma e, se lo sforzo raggiunge un valore sufficientemente elevato, si frattura. Le esperienze di laboratorio hanno dimostrato che prima della frattura avvengono mutamenti osservabili nelle proprietà fisiche della roccia, simili a quelli che si osservano talvolta in natura prima dei terremoti. Sono stati formulati modelli teorici che spiegano alcuni fenomeni precursori, mettendoli in relazione alle condizioni fisiche che determinano un terremoto. Tuttavia nessuno di questi modelli e nessuna delle possibili osservazioni e misurazioni possono essere usati allo stato attuale per formulare uno schema operativo di previsione.

Negli anni settanta, l'osservazione degli stessi fenomeni precursori in aree diverse del mondo e la previsione riuscita di alcuni terremoti hanno fatto sperare che un tale schema operativo fosse acquisibile in un tempo relativamente breve. Oggi tale obiettivo appare invece assai più lontano. I programmi di osservazione dei fenomeni precursori, realizzati negli ultimi anni in alcune aree sismiche del mondo, non hanno infatti fornito i risultati sperati, nel senso che il legame tra fenomeni precursori e terremoto si è rivelato assai più variabile e incerto di quanto si pensasse. D'altra parte la ricerca teorica non è ancora in grado di interpretare in un quadro unico tutti i fenomeni osservati. Che cosa si può fare dunque nel frattempo?

Il terremoto che ha colpito il Messico il 19 settembre 1985 è stato un effetto del moto relativo della zolla nordamericana e della zolla delle Cocos. La sua magnitudo, calcolata dall'ampiezza delle onde sismiche superficiali, è stata 8,1. Per avere un'idea dell'energia sismica liberata, essa è stata circa 90 volte superiore a quella del sisma che colpì l'Irpinia nel 1980, la cui magnitudo fu 6,8. Il 21 settembre, il terremoto è stato seguito da una forte replica, di magnitudo 7,5. La fotografia, un'immagine della capitale messicana, è di Alain Nogues (Sygma/Grazia Neri).

Si parla a volte di previsione «statistica» dei terremoti, intendendo con ciò la determinazione della «probabilità» che un terremoto di una certa magnitudo avvenga in una data zona entro un certo periodo di tempo. Il ricorso a tecniche statistiche nello studio della sismicità è dovuto alla grande complessità del fenomeno. Il sistema fisico in cui il terremoto si manifesta, la litosfera terrestre, ha proprietà meccaniche che variano nello spazio e nel tempo in maniera complicata ed è sottoposto a forze interne ed esterne altrettanto variabili e poco note. Il metodo statistico mira a comprendere e a prevedere il comportamento di un tale sistema sulla base di un numero minimo di concetti fondamentali. In particolare, ricerca i parametri più utili per descrivere i fenomeni; cerca di mettere in evidenza le caratteristiche essenziali e le regolarità che vi compaiono e di fornire metodi relativamente semplici per effettuare previsioni di tipo probabilistico. Le principali regolarità osservate consistono nell'esistenza di cicli sismici, nella presenza di lacune nell'attività sismica sia spaziali sia temporali e nella migrazione degli epicentri dei terremoti. I modelli fisici indicano che tali fenomeni sono strettamente connessi l'uno all'altro, essendo tutti dipendenti dalle condizioni di sforzo esistenti nella litosfera.

Un requisito fondamentale per affrontare statisticamente il problema della previsione è costituito dalla conoscenza del comportamento a lungo termine delle faglie attive. La durata del periodo che intercorre tra due grandi terremoti, la regolarità di quel periodo e la data dell'evento più recente ci aiutano a stabilire se siamo vicini a un altro terremoto oppure no. Purtroppo, nella maggior parte delle regioni sismiche, il periodo di documentazione storica completa è troppo breve per determinare il comportamento sismico a lungo termine di tali regioni. La documentazione storica indica che i grandi terremoti sono eventi che si ripetono. Per esempio, un segmento lungo 200 chilometri della Fossa di Nankai (che, al largo del Giappone, segna il confine tra la zolla asiatica e la zolla delle Filippine) ha prodotto terremoti negli anni 684, 887, 1099, 1361, 1605, 1707, 1854 e 1946. L'intervallo di ricorrenza varia da un minimo di 92 a un massimo di 262 anni. Questa osservazione elementare ha condotto a definire il concetto di «periodo di ricorrenza medio», che nel caso illustrato è di 180 anni. Si noti che, se si assume che la documentazione sia completa, i singoli intervalli di ricorrenza si discostano dalla media fino al 45 per cento. In base a questi dati, possiamo supporre che non avverrà un altro grande terremoto su quel segmento prima che sia trascorso almeno

Per ogni tipo di previsione dei terremoti - statistica, deterministica, a breve, medio, lungo termine - vi è un certo grado di indeterminazione nello spazio e nel tempo, qui indicato dalle dimensioni dei rettangoli. La previsione statistica dei grandi terremoti consente in genere una localizzazione spaziale migliore di quella temporale. Per la previsione deterministica, ci si attende che l'indeterminazione sia tanto più grande quanto maggiore è l'anticipo con cui la previsione viene formulata.

La previsione «statistica» dei terremoti si basa sulla conoscenza del comportamento a lungo termine delle faglie attive. I due modelli riproducono in modo semplificato il processo di accumulo e di liberazione di sforzo su un segmento di margine tettonico. In quello «a tempo prevedibile» (*a sinistra*), si assume che ciascun terremoto avvenga ogniqualvolta lo sforzo raggiunge un valore di soglia fisso: in tal caso il tempo che deve trascorrere prima del terremoto successivo è proporzionale alla caduta di sforzo. Nel modello «a scorrimento prevedibile» (*a destra*), si assume che ciascun terremoto produca una caduta di sforzo a un livello fisso: in questo caso, lo scorrimento connesso a ciascun terremoto risulta proporzionale al tempo trascorso dal terremoto precedente.

mezzo secolo da oggi. Tuttavia, la notevole variazione degli intervalli di ricorrenza non consente una previsione più precisa della data del prossimo terremoto. È inoltre scarsamente noto in che misura i terremoti che si ripetono su una stessa faglia siano simili l'uno all'altro dal punto di vista dei parametri fisici (ampiezza dello scorrimento, area della superficie dislocata, durata ecc.). I pochi esempi adeguatamente documentati mostrano che terremoti successivi possono essere notevolmente simili, ma anche notevolmente diversi.

Inoltre, l'ipotesi in base alla quale i terremoti su una data faglia hanno intervalli di ricorrenza che si dispongono attorno a un periodo di ricorrenza medio può non essere valido in certe regioni. Se in una regione sono presenti diverse faglie, con dimensioni e grado di attività simili, l'intervallo di ricorrenza può essere determinato da processi di applicazione, liberazione e ridistribuzione dello sforzo assai più complessi di quelli che avvengono lungo margini tettonici «semplici», caratterizzati da una faglia principale dominante. In una tale regione, le deviazioni dei singoli intervalli di ricorrenza dal periodo medio calcolato possono essere così grandi da rendere la stima di quest'ultimo di scarsa utilità ai fini della previsione.

La conoscenza delle date dei grandi terremoti avvenuti nel passato rimane in ogni caso un elemento importante per lo studio di una regione sismica. I documenti storici ci permettono di risalire all'indietro di almeno 2000 anni nelle regioni di più antica civiltà, anche se con problemi di completezza e di interpretazione delle fonti, ma solo di un paio di secoli nelle regioni di civilizzazione recente come la California. Si sta pertanto sviluppando una disciplina, la «paleosismologia», che ricerca e studia i terremoti preistorici. Le tecniche della paleosismologia sono in prevalenza geologiche e si applicano alle faglie che affiorano in superficie.

È chiaro che il metodo statistico, pur presupponendo l'esistenza di connessioni deterministiche nei fenomeni studiati, nulla ci può dire sulla natura dei processi fisici che sono all'origine dei terremoti. Il suo ruolo è tuttavia importante perché ci permette di definire il «rischio sismico» di una data regione. Questa informazione è estremamente utile per due aspetti: in primo luogo, per stabilire le priorità negli interventi di prevenzione dei danni; in secondo luogo, per individuare le aree nelle quali avviare programmi di osservazioni e di rilevamenti finalizzati alla previsione.

La previsione statistica appare dunque come la premessa alla vera previsione di tipo deterministico. Una previsione è definita «deterministica» se basata su leggi fisiche deterministiche. È possibile raggiungere questo obiettivo? Come si è detto, un primo tentativo è stato compiuto correlando il terremoto con i fenomeni precursori tramite leggi fenomenologiche. È il metodo utilizzato spesso nelle fasi iniziali di una scienza, quando non sono ancora disponibili modelli e teorie per interpretare i fenomeni. Per esempio, se si riuscisse a stabilire che un terremoto di una certa magnitudo è sempre preceduto da un fenomeno precursore di una certa entità misurabile, si otterrebbe una legge fenomenologica non diversa dalle leggi sui pianeti di Keplero o da quelle di Galileo sul moto dei gravi. Il livello superiore di descrizione è quello di formulare modelli e teorie che abbiano un valore esplicativo sempre più ampio (come la teoria di Newton, quella di Einstein e via dicendo, per restare all'esempio della gravitazione). Il problema in sismologia è che la complessità del sistema studiato non ha permesso finora di formulare tali leggi fenomenologiche. Forse non lo consentirà neppure in futuro. Ciò tuttavia non esclude in linea di principio la possibilità di una previsione deterministica.

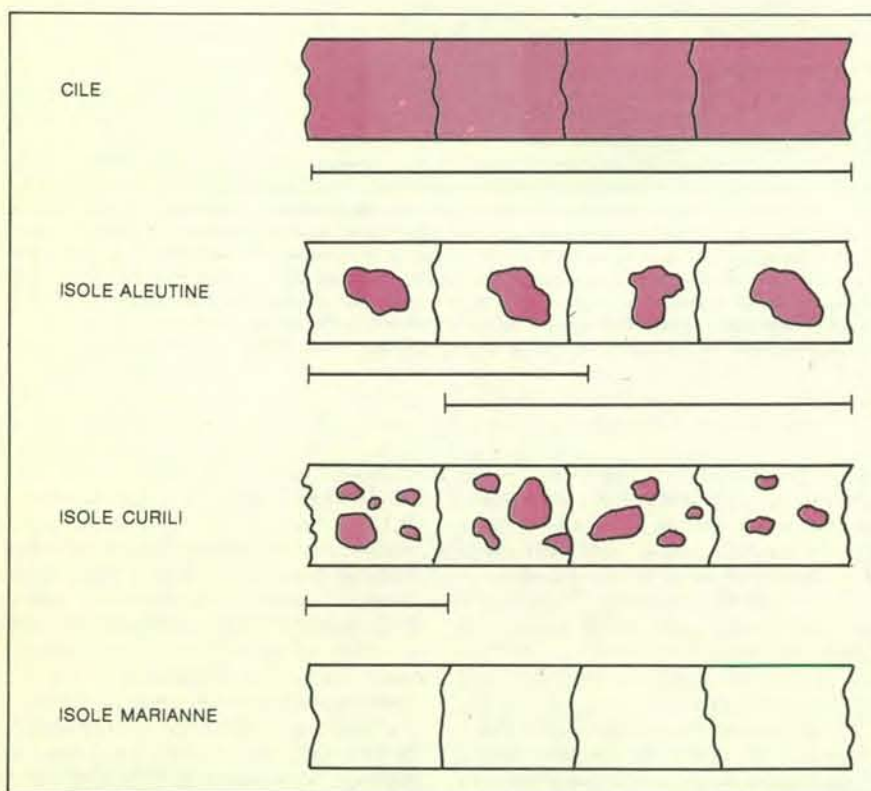
Il metodo empirico per affrontare il problema della previsione, cioè il tentativo di formulare leggi di tipo fenomenologico, ha condotto numerosi ricercatori a proporre relazioni che collegano i fenomeni precursori al terremoto. Citiamo a titolo di esempio le leggi che mettono in relazione il tempo di comparsa dei fenomeni precursori e la magnitudo del terremoto che essi precedono. Sono state proposte leggi diverse a seconda dei dati presi in considerazione, anche se quasi tutte stabiliscono una relazione secondo cui l'anticipo del precursore aumenta in maniera esponenziale con la magnitudo del terremoto. Nessuna di queste leggi tuttavia ha un sufficiente carattere di precisione e generalità.

Queste difficoltà non devono sorprenderci. I fenomeni precursori sono stati osservati spesso quasi casualmente e non tramite un programma di osservazioni sistematiche durate secoli, come nel caso dei moti celesti. La mancanza di osservazioni a lungo termine effettuate presso le faglie attive ha portato probabilmente a sottolineare gli esempi di correlazione positiva tra un dato precursore e il terremoto successivo. A causa della brevità del periodo di osservazione e della rarità dei grandi terremoti sulla stessa faglia, è impossibile stabilire con una certa sicurezza se un presunto precursore è veramente tale. Esistono anche osservazioni di fenomeni anomali, che potrebbero essere interpretati come precursori, ma non sono stati seguiti da terremoti chiaramente identificabili in base a essi.

Dal punto di vista fisico, una volta individuata la faglia sorgente di un possibile terremoto, il problema è sostanzialmente quello di determinare le condizioni di instabilità che ne producono lo scorrimento. Una volta compreso il processo di caricamento e di liberazione dello sforzo dal punto di vista teorico, si tratta di misurare i parametri che caratterizzano la stabilità o l'instabilità della faglia. Da ciò possiamo forse comprendere perché il metodo puramente empirico fallisce. Vi è una dispersione delle osservazioni su fenomeni secondari, mentre occorre concentrare l'attenzione su quei parametri i cui cambiamenti nel tempo forniscono informazioni dirette sui processi fisici che portano allo scorrimento della faglia. In base agli attuali modelli della sorgente sismica, questi parametri sono la deformazione del suolo a scala regionale e a scala locale lungo la faglia e la sismicità minore sulla faglia stessa e su quelle adiacenti. Sono questi i fenomeni che ci informano direttamente sullo stato di sforzo nella regione.

Si può sostenere che i campi elettrici e magnetici, la velocità delle onde sismiche, l'emanazione di gas dal suolo o il comportamento degli animali possono costituire indicatori più sensibili dell'accumulo di sforzo in una regione. Ma esistono poche osservazioni verificabili che confermino questa assunzione. Inoltre non è chiaro come si possa risalire dalle variazioni di sforzo dedotte da tali osservazioni al grado di instabilità della faglia. Per quanto riguarda le emissioni di gas rado, sono state osservate anomalie interessanti, ma non esiste alcun modello che spieghi in che modo usare tali anomalie per prevedere i terremoti. Anche il livello dell'acqua nei pozzi dà informazioni sulla deformazione del suolo, ma in definitiva fornisce scarsi elementi riguardo al processo di scorrimento che produce il terremoto. Ci troviamo di fronte troppi fenomeni che, non spiegati da modelli e non direttamente connessi al processo fondamentale che dà origine al terremoto, possono condurci a conclusioni errate e a falsi allarmi. Laddove si intenda realizzare un programma di osservazione e di controllo volto allo studio di una data area sismica con l'obiettivo della previsio-

Nel tentativo di affrontare con metodo empirico il problema della previsione dei terremoti sono state suggerite da parte di studiosi diversi alcune relazioni che collegano il tempo di comparsa dei fenomeni precursori e la magnitudo del terremoto che essi precedono. I precursori compaiono in genere con un anticipo che cresce esponenzialmente all'aumentare della magnitudo del terremoto. (Fanno eccezione i fenomeni precursori che seguono la relazione rappresentata dalla curva in basso, per i quali il tempo di comparsa è pressoché uniforme per magnitudo maggiori di 5.) A causa del rapido aumento del tempo di comparsa con la magnitudo, si usa una scala logaritmica, in cui i tempi sono espressi in giorni precedenti il terremoto. Per consentire una comprensione più immediata, sulla destra sono riportate in corrispondenza alcune unità temporali di uso comune.

Data la particolare importanza che si attribuisce in teoria alla eterogeneità della sorgente sismica sono stati costruiti modelli per spiegarne il ruolo. Nel «modello a barriera» (*in alto*), un piano di faglia rettangolare è sottoposto a sforzo uniforme (*in colore*). Lo scorrimento connesso al terremoto libera dallo sforzo solo una parte della faglia (*area in bianco*), incontrando un ostacolo nella barriera. Nel «modello ad asperità» (*in basso*), lo sforzo è inizialmente concentrato solo su una parte della faglia, la cosiddetta «asperità» (*in colore*), il cui cedimento produce il terremoto.

Lo studio delle caratteristiche dei grandi terremoti superficiali ha permesso di classificare alcune zone sismiche in base alle dimensioni delle asperità presenti. Nell'illustrazione è riportata la distribuzione delle asperità (*in colore*) su alcuni margini di zolla attorno all'oceano Pacifico ipotizzata da Hiroo Kanamori e collaboratori del California Institute of Technology. I segmenti riportati sotto i margini del Cile, delle Aleutine e delle Curili rappresentano l'estensione dello scorrimento nei grandi terremoti. Sul margine delle isole Marianne si ha per lo più scorrimento asismico.

ne, occorre invece concentrare le misurazioni su parametri quali la deformazione e la sismicità, legati direttamente ai processi fisici che portano allo scorrimento delle faglie e ai grandi terremoti.

Deve essere chiaro che esprimere un giudizio negativo sul metodo empirico non equivale affatto a negare la possibilità di una previsione deterministica. Tutt'altro. Per esempio, se per una certa regione conoscessimo le leggi che regolano il comportamento meccanico delle rocce e lo stato di sforzo in tutti i punti della regione e in ogni istante di tempo, potremmo prevedere il prodursi di un terremoto senza bisogno dei fenomeni precursori.

Esiste però un altro problema. La previsione è il risultato dell'inserimento di dati di osservazione in una teoria. Questi dati rappresentano i valori numerici delle grandezze fisiche che compaiono nella teoria, comprese le «condizioni iniziali» del sistema studiato. Ma una teoria può essere deterministica e tuttavia la sua applicazione a circostanze specifiche può non essere in grado di fornire previsioni accurate a causa della carenza di conoscenza empirica. Per esempio, una piccola imprecisione nelle condizioni iniziali può alterare completamente, dopo breve tempo, l'evoluzione del sistema. È questo il caso della meccanica dei fluidi applicata alla previsione meteorologica. Ed è anche il caso della teoria dell'elasticità e della meccanica delle fratture applicate allo studio dei terremoti. Perciò una imprecisione troppo elevata nelle condizioni iniziali o negli altri dati del problema oppure l'inadeguatezza del modello può portare a una situazione in cui l'evento previsto non avviene.

Si può anche sostenere un punto di vista più radicale, cioè che non è possibile giungere alla previsione deterministica, perché non si può costruire una teoria deterministica dei fenomeni sismici. È in effetti possibile delineare modelli teorici che comportano l'impossibilità di una previsione sulla base di fenomeni precursori. Infatti, anche se la faglia è sottoposta a uno sforzo considerevolmente più basso di quello che provoca la nucleazione di un grande terremoto, lo scorrimento può essere innescato da uno o più piccoli terremoti o da scorrimento asismico sulla stessa faglia o su una faglia vicina. Dunque, per prevedere quando avverrà uno scorrimento su una particolare sezione della faglia, è necessario prevedere quando si avrà scorrimento sulle sezioni adiacenti della stessa faglia o su faglie vicine e se questo scorrimento è in grado di innescare il terremoto. Ma lo stesso ragionamento può essere applicato alle sezioni adiacenti e alle faglie vicine e così via, fino a una presumibile nucleazione iniziale che ha prodotto la catena di eventi. Tale nucleazione avviene probabilmente a livello microscopico, laddove lo sforzo è abbastanza elevato da produrre spontaneamente lo scorrimento o da permetterne l'innesco da parte di agenti deboli come le maree solide, o la pressione atmosferica, o la graduale crescita di sforzo tettonico. In questo modello, per prevedere un grande terremoto occorre prevedere la nucleazione iniziale. Quindi è necessario prevedere tutti i piccoli terremoti e determinare quali di essi produrranno per stadi successivi un grande terremoto. È evidente l'impossibilità di un tale programma. I fenomeni precursori osservati possono essere soltanto correlati in maniera probabilistica ai terremoti, nel senso che, per un sistema complesso in condizioni di instabilità, qualunque cambiamento dei parametri ha una forte probabilità di essere precursore. Allo stato attuale della nostra comprensione dei terremoti, non è possibile scartare un modello di questo tipo. Esistono invece alcune osservazioni che possono essere portate a suo sostegno.

Il punto di vista più comune e «ottimistico» è tuttavia quello di formulare modelli nel cui ambito i terremoti possono essere previsti e di confrontare tali modelli con le osservazioni per valutarne l'adeguatezza. Notevoli passi in avanti sono stati compiuti negli ultimi anni dal punto di vista teorico. In particolare, si è compreso che un ruolo molto importante hanno le eterogeneità della sorgente sismica. I termini «asperità» e «barriera» sono stati introdotti per definire due tipi caratteristici di eterogeneità sulle superfici di faglia. Entrambi i termini indicano parti della superficie di faglia che presentano una resistenza elevata allo scorrimento dei due lembi della faglia, cioè un attrito elevato. Essi vengono però usati per indicare due ruoli diversi svolti da queste parti più resistenti.

Nel «modello a barriera» si assume che lo sforzo sulla faglia sia uniforme, mentre l'attrito è variabile, essendo più alto nella barriera. Può dunque accadere che solo una parte della superficie di faglia scorra durante il terremoto, mentre la barriera si oppone allo scorrimento. In questo modo però sulla barriera viene concentrato ulteriore sforzo. Di conseguenza, dopo il terremoto, la barriera può essere erosa da episodi di scorrimento asismico oppure da piccole scosse successive (le «repliche» del terremoto).

Nel «modello ad asperità» si assume che lo sforzo sulla faglia sia variabile (maggiore in corrispondenza dell'asperità), mentre l'attrito è considerato uniforme. Una situazione del genere può verificarsi se scorrimento asismico e scosse preliminari hanno liberato dallo sforzo la parte di faglia circostante l'asperità, concentrandolo su questa ultima. Quando lo sforzo raggiunge un valore critico, l'asperità cede, dando luogo al terremoto.

I due modelli rappresentano situazioni limite. In generale, su una faglia esisteranno sia asperità sia barriere, cioè vi saranno eterogeneità sia nello sforzo sia nell'attrito. Inoltre una stessa porzione di faglia può passare da un ruolo all'altro, comportandosi da barriera in un terremoto e da asperità in un terremoto successivo. Lo studio delle caratteristiche dei grandi terremoti superficiali che avvengono lungo le coste dell'oceano Pacifico ha permesso di classificare le diverse zone sismiche in base alle dimensioni delle asperità presenti. Se la distribuzione delle asperità che contrassegna una data faglia è una caratteristica che si mantiene almeno per un certo numero di cicli sismici, la conoscenza della «mappa» delle asperità sarà un elemento fondamentale per un programma di previsione. Ogni zona sismica ha avuto la sua storia e avrà la sua evoluzione: occorre studiare zona per zona.

La non uniformità delle proprietà meccaniche e delle condizioni di sforzo lungo i margini delle zolle è evidenziata tra l'altro dal fatto che lo scorrimento connesso ai grandi terremoti riguarda solo un segmento del margine stesso. Poiché lo scorrimento su un segmento concentra sforzo oltre i suoi estremi, terremoti successivi avvengono su segmenti diversi, fino a coprire totalmente il margine dopo un tempo sufficientemente lungo (certi tratti del margine possono però scorrere in maniera asismica). La distribuzione spaziale e temporale degli eventi sismici lungo il margine è complessa e varia da una regione all'altra, ma consente spesso di individuare segmenti che in un certo ciclo sismico non sono stati ancora interessati da terremoti, mentre i tratti adiacenti hanno già subito scorrimento: si tratta delle «lacune sismiche», dove la probabilità di un futuro terremoto è massima. Diversi dei grandi terremoti più recenti sono avvenuti in aree individuate come lacune.

Sottolineiamo dunque come la disponibilità di una teoria adeguata dei fenomeni sismici non sia sufficiente per effettuare una previsione deterministica. Per ottenere previsioni in situazioni reali, la teoria ha necessità di dati: le proprietà meccaniche delle rocce e lo stato di sforzo della litosfera nella regione interessata devono essere conosciuti in modo molto più approfondi-

Anche se lungo un margine di zolla la distribuzione spaziale e temporale degli eventi sismici è molto complessa, è possibile a volte individuare segmenti che, in un certo ciclo sismico, non sono stati ancora interessati da grandi terremoti («lacune sismiche»), adiacenti ad altri che hanno già subito scorrimento. Ne è un esempio il margine tra la zolla nordamericana e la zolla delle Cocos. Nella illustrazione le aree in colore rappresentano le porzioni del margine che hanno subito scorrimento in occasione dei grandi terremoti registrati nei primi 80 anni del nostro secolo e sono state tracciate in base alla distribuzione delle repliche di ogni terremoto. Nella lacuna sismica del Michoacan, dove non avvenivano grandi terremoti dal 1911, vi è stato quello disastroso del 19 settembre 1985.

La necessità di comunicare tempestivamente alle autorità e alla popolazione il crearsi di una situazione di imminente pericolo richiede un insieme di attrezzature molto sofisticate e personale specializzato. Nella fotografia, la sala dei sismografi dell'Istituto nazionale di geofisica di Roma dove pervengono, in tempo reale e ventiquattro ore su ventiquattro, i dati raccolti da circa 50 stazioni sismometriche distribuite su tutto il territorio nazionale. I dati vengono elaborati da un calcolatore e, in caso di necessità, comunicati immediatamente al Ministero della protezione civile.

to di quanto non lo siano attualmente.

Gli studi di paleosismologia hanno recentemente condotto a formulare il modello del «terremoto caratteristico». Si tratta di un terremoto che si ripete periodicamente su un certo segmento di faglia, con lo stesso meccanismo e la stessa magnitudo. Un esempio di terremoto caratteristico è stato osservato su un segmento della faglia di San Andreas, presso la città di Parkfield. Il segmento, lungo una ventina di chilometri, corrisponde alla zona di transizione tra due comportamenti radicalmente diversi della faglia. A nord di questo segmento, lo scorrimento avviene in maniera asismica e continua, accompagnato da piccoli terremoti (magnitudo minore di 4); a sud, la faglia è invece «bloccata» e scorre soltanto in occasione di grandi terremoti (magnitudo maggiore di 7), l'ultimo dei quali è stato il terremoto di Fort Tejon del 1857. Dal 1857 a oggi, sul segmento di Parkfield sono avvenuti cinque terremoti di magnitudo attorno a 6, con caratteristiche assai simili. Il periodo di ricorrenza medio è di circa 22 anni, con una deviazione standard di appena tre anni. L'ultimo terremoto della serie è avvenuto nel 1966. Queste osservazioni, interpretate in base ai modelli ad asperità e a barriera, hanno dato il via a un interessante «esperimento» di previsione. Il prossimo terremoto dovrebbe avvenire nel 1988, con una indeterminazione di alcuni anni in più o in meno. Una rete di strumenti è stata installata nell'area interessata. Si spera in questo modo di apprendere qualche cosa di più sulle variazioni dei parametri fisici che si manifestano prima dei terremoti.

La notevole indeterminazione connessa a qualunque tentativo di previsione oggi immaginabile pone seri problemi nel caso in cui tale previsione venga comunicata alla popolazione dalle istituzioni scientifiche o dagli enti preposti alla protezione civile. Diversamente da altre manifestazioni violente della natura, come inondazioni, uragani e trombe d'aria, i terremoti non sono preceduti normalmente da segnali visibili a tutti e in base ai quali la popolazione possa identificare l'imminenza dell'evento e ottenere conferma della previsione. Se una previsione su base scientifica indica un intervallo temporale di un giorno, un mese o un anno, il singolo individuo non è in grado di osservare segnali che gli dicano quando lasciare la propria abitazione, se necessario, o prendere altre precauzioni dell'ultimo minuto, come invece avverrebbe nel caso di altri fenomeni naturali. D'altra parte se un terremoto previsto dai sismologi non avviene, il pubblico non si renderà conto del fatto che esso sarebbe potuto avvenire o che potrà avvenire in seguito. Per i sismologi è chiaro che, se si accumula sforzo in una regione, alla fine dovrà aver luogo un terremoto: diversamente da un uragano o da un tornado, il terremoto non può andare da un'altra parte. Ma non vi saranno segni, evidenti alla popolazione, che confermino la convinzione dei sismologi che la probabilità di un terremoto è aumentata, e non diminuita, se esso non è avvenuto quando era stato previsto. Nel caso di previsione errata, il sismologo si trova analogamente in una posizione delicata, perché non è in grado di ricorrere a nient'altro che alle sue stesse osservazioni per giustificare l'errore. Queste circostanze distinguono i terremoti dagli altri fenomeni naturali distruttivi, nel senso che la reazione della popolazione dipende esclusivamente dalla fiducia che intende riporre nelle previsioni scientifiche. La preparazione e il senso di responsabilità delle popolazioni coinvolte sono pertanto elementi essenziali per la riuscita di esperimenti di previsione resi pubblici, sia nel caso che la previsione abbia successo sia in caso contrario.

L'arte muraria degli inca

Eseguendo personalmente il lavoro di tagliapietre in una cava incaica, l'autore è riuscito a spiegare come gli inca raggiungessero tanta precisione nel far combaciare gli enormi blocchi di pietra usati nelle loro costruzioni

di Jean-Pierre Protzen

La cultura incaica raggiunse l'apogeo solo un centinaio di anni prima della conquista spagnola, avvenuta nel 1532. Durante quel secolo la società incaica si trasformò da piccolo stato agricolo del Perù centrale a un potente impero che si estendeva dal Cile all'Ecuador. Un aspetto della sua fioritura culturale fu un ambizioso programma di nuove costruzioni iniziato nel 1438 da Pachacutec, il nono *inca* (o imperatore). Pachacutec ordinò alle sue maestranze di ricostruire Cuzco, la capitale dell'impero emergente. La ricostruzione non si fermò alla morte del nono inca. I suoi successori estesero le nuove costruzioni molto oltre i confini di Cuzco. In tutto il Perù furono eretti templi, palazzi, magazzini e impianti idrici, opere costruite *ex novo* o in sostituzione di strutture precedenti.

Il programma di costruzioni di Pachacutec fu non solo ambizioso, ma anche tecnicamente innovatore. Benché la maggior parte delle strutture incaiche anteriori fosse costruita probabilmente con adobe o con pietre cementate con fango, i nuovi edifici furono eretti per intero a secco. Blocchi di pietra, il cui peso poteva raggiungere anche 100 tonnellate, venivano squadrati e messi in opera con tanta precisione che in molte delle commessure non è possibile infilare neppure la lama di un coltello.

Per secoli e secoli i visitatori del Perù sono stati affascinati dalla mole dei conci usati nelle costruzioni in pietra degli inca e dalla precisione con cui ogni concio è stato inserito nell'opera muraria. Il fatto che gli inca non possedessero utensili in ferro rende queste costruzioni ancora più impressionanti. Nel 1979, mentre stavo tornando negli Stati Uniti dopo un incarico didattico temporaneo in Brasile, visitai alcuni fra i principali siti incaici e rimasi stupito dall'ingegnosità delle costruzioni. Quando chiesi alle mie guide in che modo gli inca squadrassero le grandi pietre e le componessero negli edifici ricevetti risposte non del tutto soddisfacenti. Tornato all'Università della California a Berkeley, dove insegno architettura, chiesi a colleghi archeologi informazioni bibliografiche sulle costruzioni in pietra degli inca. Con mia sorpresa, mi fu detto che su quest'argomento non esisteva niente.

Pur non essendo un archeologo, ho un acuto interesse professionale per le tecniche di costruzione. Dopo avere riflettuto a lungo sull'argomento, decisi di compiere personalmente ricerche sulle opere in muratura degli inca. Un semestre di congedo per studi, nel 1982, mi diede l'opportunità di trascorrere sei mesi in Perù, e da allora vi sono tornato ogni anno per un mese circa. La mia ricerca non si è fermata alla fase delle ipotesi. Una volta formulata un'ipotesi, la sottoponevo immediatamente a verifica. Usando i materiali disponibili nei siti incaici, sgrossai, squadrai e composi assieme conci per dimostrare che questi compiti potevano essere eseguiti dagli inca nel modo da me ipotizzato. Rimangono alcuni misteri, in particolare su come i grandi blocchi in pietra venivano trasportati e manovrati nei siti di costruzione, ma nel complesso la mia ricerca ha avuto successo. Oggi è quindi possibile cominciare a sostituire le congetture sul modo in cui gli inca costruirono le loro belle strutture in pietra con dati empirici.

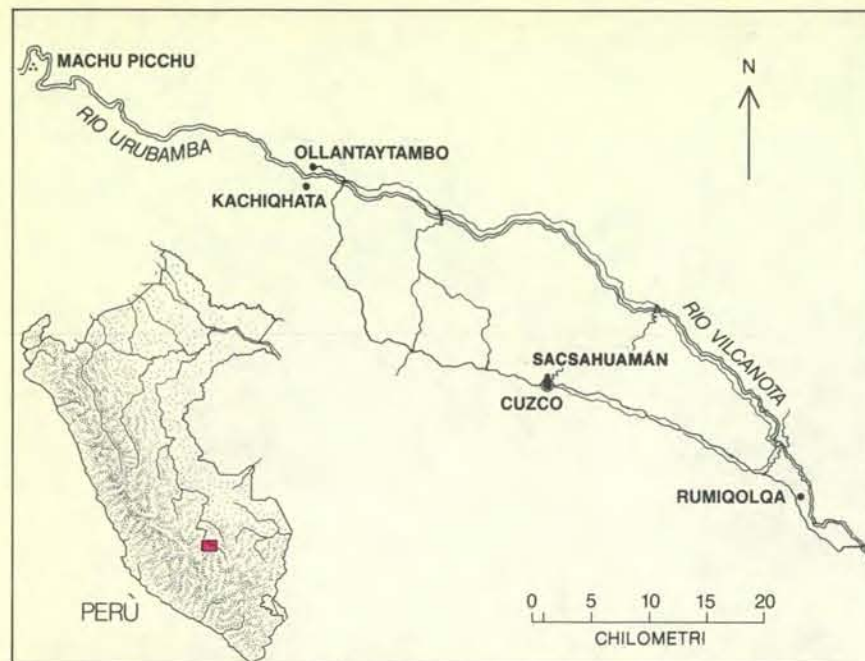
Gran parte della mia ricerca è consistita nell'analisi di specifiche mura incaiche nella stessa Cuzco e nelle «fortezze» di Sacsahuamán e di Ollantaytambo. Sacsahuamán è nei pressi di Cuzco e Ollantaytambo si trova sul Rio Urubamba, una novantina di chilometri a nord-est della capitale incaica. Benché in numerosi testi le costruzioni di Sacsahuamán e di Ollantaytambo vengano descritte come fortezze, ricerche archeologiche recenti inducono a pensare che avessero una funzione religiosa e non militare. Ma quale che sia stato il ruolo che quelle fortezze svolsero nella società incaica, i due siti sono impressionanti dal punto di vista delle tecniche di costruzione. Sacsahuamán è un sito molto grande, comprendente tre terrazze sovrapposte, cinte da tre mura in pietra separate, alte più di tre metri. Ollantaytambo, costruito sul contrafforte di una montagna, comprendeva un centro religioso, una proprietà reale e una città costruita secondo un piano urbanistico.

Come venivano costruite queste grandi strutture in pietra? Per poter meglio trattare questo problema, l'ho suddiviso in quattro parti: l'estrazione delle pietre dalle cave, la sgrossatura e la squadratura dei singoli blocchi, la rifinitura dei conci in modo da farli combaciare perfettamente e il trasporto. Per la ricerca sull'estrazione dei blocchi ho visitato varie cave incaiche, analizzandone due particolareggiatamente, Kachiqhata e Rumiqolqa. Kachiqhata si trova a quattro chilometri circa da Ollantaytambo, al di là del Rio Urubamba, e le sue cave avevano fornito il porfido (granito rosso) usato nella costruzione del Tempio del Sole, la struttura più importante a Ollantaytambo. Rumiqolqa si trova 35 chilometri a sud-ovest di Cuzco e ha fornito la maggior parte dell'andesite (una roccia vulcanica) usata dai muratori di Pachacutec nella ricostruzione della capitale imperiale.

Vari indizi suggeriscono che l'estrazione di pietre da costruzione fosse un'attività molto importante per gli inca. Kachiqhata e Rumiqolqa sono località remote, difficili da raggiungere e lontane dai siti nei quali i blocchi venivano usati per le costruzioni. La ragione che indusse gli inca a sfruttare cave così poco accessibili deve essere stata il grande valore che essi attribuivano al tipo di pietra che vi si trova.

L'organizzazione interna delle cave dimostra inoltre quanta attenzione venisse prestata al sistema adottato per procurarsi le pietre da costruzione. Tanto Rumiqolqa quanto Kachiqhata hanno reti di vie d'accesso che conducono ai punti dai quali si estraevano le pietre da costruzione. Si arriva, per esempio, alle cave di Kachiqhata percorrendo una strada che scende da Ollantaytambo, attraversa il Rio Urubamba e sale, sull'altra riva del fiume, fino a una serie di dirupi dai quali si staccano naturalmente falde di roccia che si accumulano

Sacsahuamán, nei pressi di Cuzco, è il sito di una delle più imponenti costruzioni in pietra degli inca. La fotografia mostra parte del sistema di tre terrazze sovrapposte comprendente tre cerchie di mura alte più di tre metri. I massi più grandi pesano circa 100 tonnellate. Benché Sacsahuamán venga spesso descritta come una fortezza, ricerche archeologiche recenti fanno pensare che si trattasse probabilmente di un centro religioso.

Nel particolare di un muro di Ollantaytambo è evidente la notevole precisione con cui i conci delle costruzioni incaiche combaciano. Il materiale dei blocchi è meta-arcose, un tipo di arenaria. Le sporgenze servivano a manovrare i massi nel sito di costruzione; spesso, a lavoro ultimato, erano lasciate sui blocchi. Questi erano coperti di piccoli segni di scheggiatura prodotti dalle pietre usate come martelli per squadrare i conci. I segni di scheggiatura sono più minuti ai bordi che nel centro della faccia, il che fa pensare all'uso di percussori diversi nelle due aree.

Alcuni siti incaici sono concentrati nei pressi di Cuzco, negli altopiani del Perù centromeridionale. Cuzco era la capitale dell'impero degli inca, e ivi la tecnica di costruzione a secco con pietre perfettamente squadrate raggiunse un nuovo culmine nel Quattrocento. Le pietre usate nella costruzione di molti edifici provenivano dalle cave di Rumiqolqa. A Ollantaytambo sorge un'imponente rovina incaica che, come la costruzione di Sacsahuamán, è spesso designata come fortezza, mentre era probabilmente un centro religioso. Le pietre usate a Ollantaytambo provenivano da Kachiqhata. Machu Picchu, una delle città incaiche più famose e meglio situate, sorge fra picchi montuosi. Urubamba e Vilcanota sono due nomi di uno stesso fiume in parti diverse del corso.

a valle in grandi pile. Là dove raggiunge questi dirupi, la via di accesso si suddivide in numerose ramificazioni che conducono ai vari siti della cava. Il percorso della strada può essere ricostruito con facilità in quanto il fondo è ragionevolmente ben conservato ed è inoltre fiancheggiato da un'ottantina di blocchi di pietra che furono abbandonati dagli inca.

Nella parte alta della collina, sul pianoro e lungo i pendii meno scoscesi della parte bassa della collina, la rete di vie d'accesso era formata da rampe che in origine erano probabilmente ricoperte da ghiaia. Sui pendii ripidi della parte bassa della collina, le rampe sono sostituite da piani inclinati lungo i quali si lasciava che i blocchi scivolassero liberamente. A Kachiqhata il piano inclinato più lungo ha una pendenza impressionante di 40 gradi, con un dislivello verticale di 250 metri; al fondo del pendio vi sono quattro blocchi abbandonati. Le cave di Rumiqolqa sono state molto sfruttate dopo la conquista e non sono ben conservate come quelle di Kachiqhata, ma anche in esse si può ricostruire una rete di strade che conducono ai siti delle cave. In entrambe le località gli inca avevano integrato le vie di accesso con altre strutture, come muri di sostegno, canali di scolo e acquartieramenti.

Benché le due località abbiano una pianta simile, i metodi usati nelle cave di Kachiqhata erano leggermente diversi da quelli usati a Rumiqolqa. A Kachiqhata gli inca non estraevano pietre nel senso tecnico, che implica il taglio della pietra da una parete di roccia o il suo distacco dalla roccia in posto mediante taglio alla base. I cavatori si limitavano invece a ispezionare i materiali rocciosi franati e a scegliere i blocchi di granito rosso a grana grossa che corrispondevano alle loro esigenze. Le osservazioni da me compiute suggeriscono che, dopo essere stato scelto a Kachiqhata, un blocco venisse lavorato assai poco prima di essere trasportato a Ollantaytambo. Pare che l'ulteriore lavorazione di squadratura della pietra e la rifinitura per far combaciare perfettamente le superfici dei vari conci venissero eseguite nel sito di costruzione.

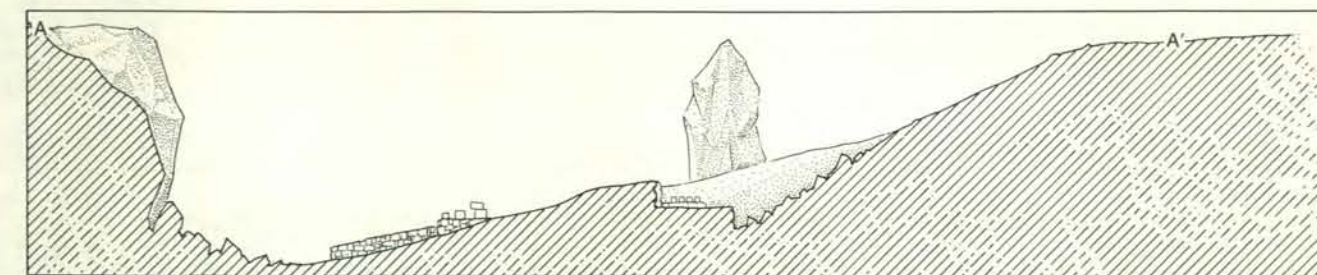
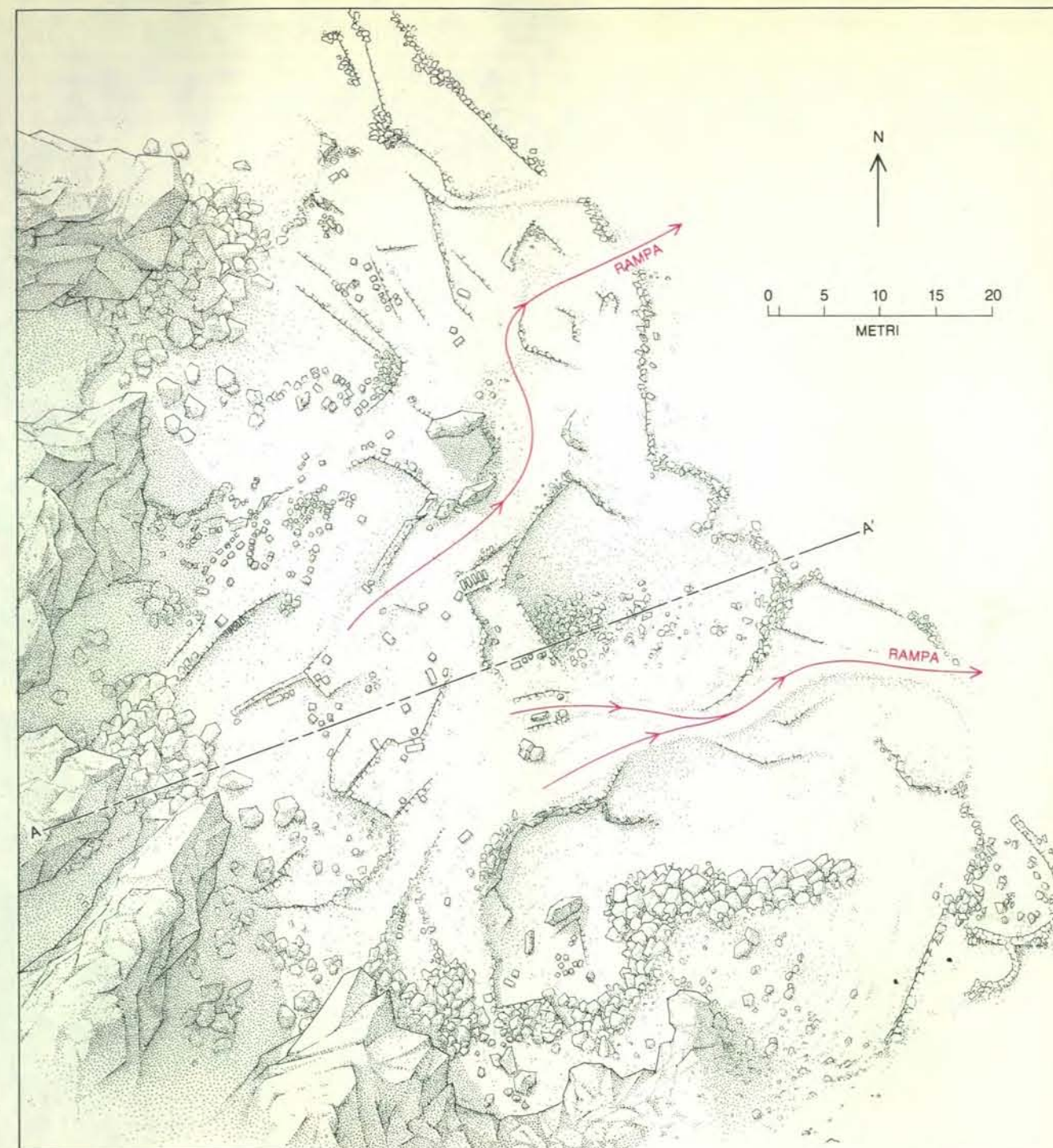
A Kachiqhata il lavoro di sgrossatura su un blocco veniva spesso cominciato prima che la rampa che conduceva a esso fosse completata. Che fosse così risulta particolarmente chiaro alla fine della rampa più alta nella cava meridionale. Nei suoi pressi, due blocchi enormi, uno di 4,5 × 2,5 × 1,7 metri e l'altro, invece, di 6,5 × 2,7 × 2,1 metri, sono sollevati su piattaforme per la lavorazione della pietra. Benché i blocchi siano parzialmente squadrati, la rampa di accesso non si estende sino alle piattaforme su cui essi si trovano.

Curiosamente, i segni di taglio riscontrati su questi blocchi e su altri che sono stati rinvenuti nelle cave incaiche sono molto simili a quelli che si possono osservare sul *pyramidion* dell'obelisco non terminato di Assuan, in Egitto. (Il *pyramidion* è la piccola punta piramidale alla sommità di un obelisco.) Tanto il pyramidion di Assuan quanto le pietre di Kachiqhata presentano, in superficie, degli incavi. È noto che gli egizi davano alle pietre la forma voluta martellandole con palle di dolerite (una roccia vulcanica). Sembra ragionevole pensare che anche gli inca facessero altrettanto.

Nel corso di un'accurata perlustrazione sul terreno della cava di Kachiqhata ho trovato alcune pietre arrotondate di quarzite, un'arenaria metamorfica che non si trova naturalmente fra le pietre della cava, mentre è presente lungo le rive del vicino Rio Urubamba. Un esame effettuato su di esse rivelò la presenza di incavi alla loro estremità più piccola, il che indica che furono usate per martellare. Ho così concluso che i cavatori inca a Kachiqhata raccoglievano ciottoli di fiume arrotondati sulle rive del Rio Urubamba e li utilizzavano poi come martelli per sgrossare i blocchi prima di procedere all'operazione più accurata di squadratura e di rifinitura che doveva essere eseguita nel sito di costruzione a Ollantaytambo.

A Kachiqhata, quindi, si sceglievano blocchi di pietra dai materiali franati più che estrarli in senso tecnico, e si dava a essi soltanto una squadratura approssimativa prima di effettuare il trasporto. A Rumiqolqa, invece, i materiali venivano scavati veramente, ossia i blocchi venivano staccati dalla parete rocciosa. Poiché in questa località le cave sono state sfruttate dopo la conquista e vengono coltivate ancora oggi, gran parte delle prove che gli inca sfruttassero la roccia sono state cancellate. Sono riuscito però a trovare un punto di scavo ben conservato in un'area difficile da raggiungere e che perciò non è stata coltivata in tempi moderni. L'ho chiamata «Cava dei lama» con riferimento ai due petroglifi di lama, incisi su una parete di roccia.

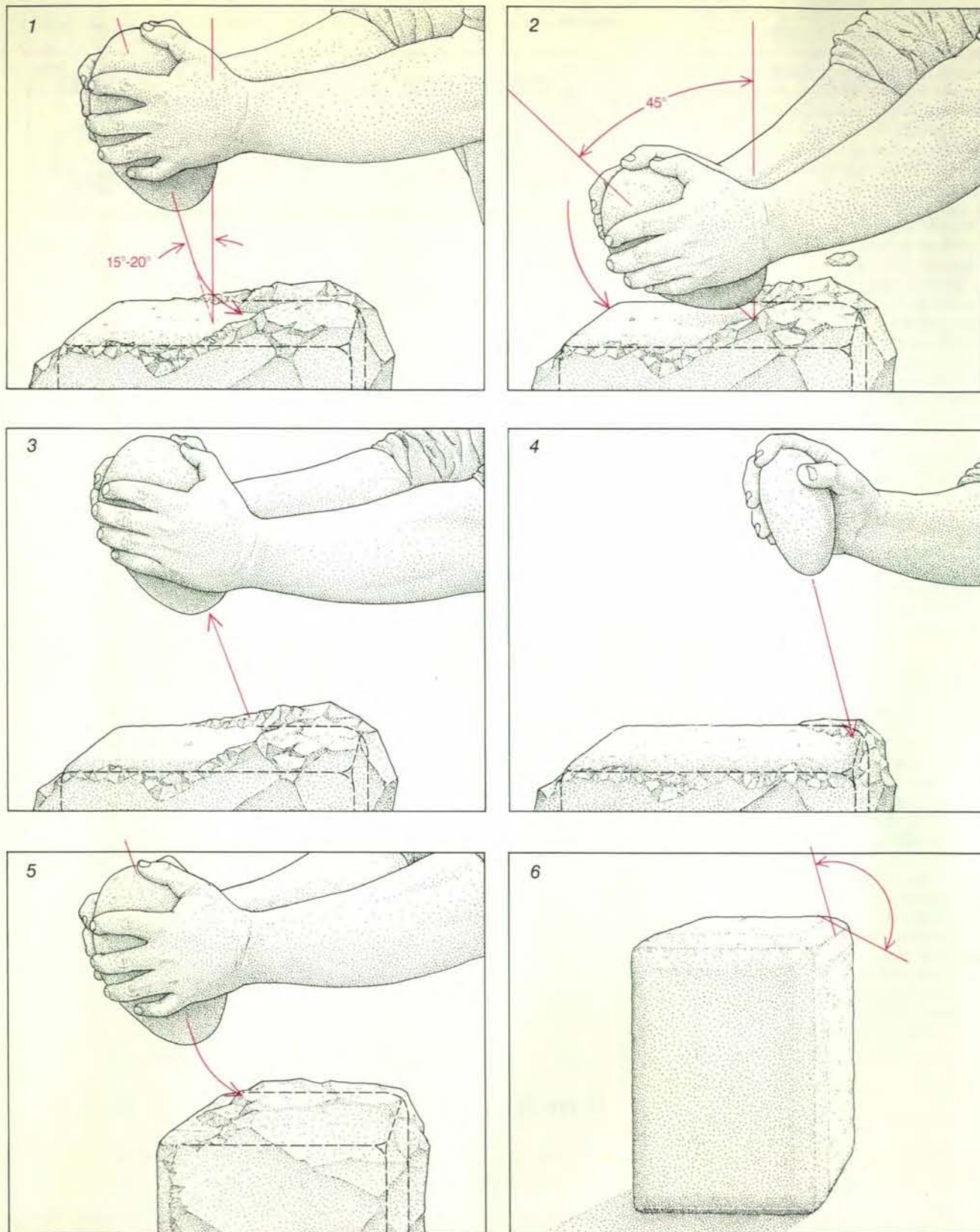
La Cava dei lama si è rivelata una ricca fonte di informazioni sul modo in cui gli inca estraevano e squadravano le pietre da costruzione. A Rumiqolqa l'estrazione dell'andesite non pone gravi problemi tecnici. Persino la roccia più densa si frattura abbastanza facilmente allo stato naturale da poterla staccare senza eccessiva difficoltà dalla superficie rocciosa. Può darsi che i cavatori staccassero dalla superficie della roccia le pietre che volevano per mezzo di leve del tipo di quelle che sono state trovate in altri siti incaici. Le leve, in bronzo, sono lunghe un metro circa; esse hanno estremi appuntiti e una sezione quadrangolare di quattro o cinque centimetri di lato. A Rumiqolqa, però, la pietra è talmente fratturata che l'uso di leve in bronzo non sarebbe stato necessario. Ho visto cavatori staccare pietre dalla parete di roccia con semplici bastoni; gli inca potrebbero aver fatto lo stesso.

Mentre è abbastanza facile capire come gli inca estraevano le pietre, è più difficile individuare le tecniche con cui le pietre venivano sgrossate e squadrate. Anche a

La Cava dei lama a Rumiqolqa è ben conservata. I cavatori incaici staccavano massi dalla parete rocciosa (*a sinistra*) servendosi probabilmente di leve di bronzo o di bastoni di legno. I blocchi da costruzione venivano poi squadrati e rifiniti prima di essere trasportati via dalla cava. Sul terreno, nella Cava dei lama, sono disseminati 250 blocchi abbandonati in varie fasi di lavorazione. Esaminandoli, l'autore ha potuto ricostruire i metodi di lavorazione della pietra usati dagli inca. Le pietre finite venivano portate via dalla cava lungo rampe ricoperte di ghiaia.

questo proposito la Cava dei lama è risultata utile. La cosa più sorprendente è costituita dai 250 blocchi di pietra estratti dagli inca e disseminati nel sito. In contrasto con le pietre di Kachiqhata, quelle di Rumiqolqa venivano in generale rifinite, o quasi rifinite, su cinque delle sei superfici quando ancora si trovavano nella cava. Fra i 250 blocchi presenti nella Cava dei lama si possono trovare esempi di tutte le fasi della lavorazione, dalla pietra grezza a blocchi squadrati con cura. Esaminando queste varie fasi sono riuscito a ricostruire il processo di lavorazione dei conci.

In primo luogo mi sono prefisso di identificare gli utensili usati per squadrare le pietre della Cava dei lama. Disseminate fra i frammenti di andesite, ho trovato pietre estranee al sito sia per forma sia per composizione. Mi sono imbattuto in un numero di queste pietre estranee sufficiente a darmi la certezza che esse venissero usate come martelli per conferire ai blocchi la forma desiderata. Come a Kachiqhata, la maggior parte di queste pietre estranee è costituita da ciottoli di fiume. Sembra che questi ciottoli provenissero dalle rive del Rio Vilcanota, che scorre non lontano dalla cava. Alcune fra le pietre usate come martelli sono di quarzite pura, altre di granito, altre ancora di basalto olivinico. (Il basalto è una roccia vulcanica e l'olivina è un minerale che si trova in esso.)

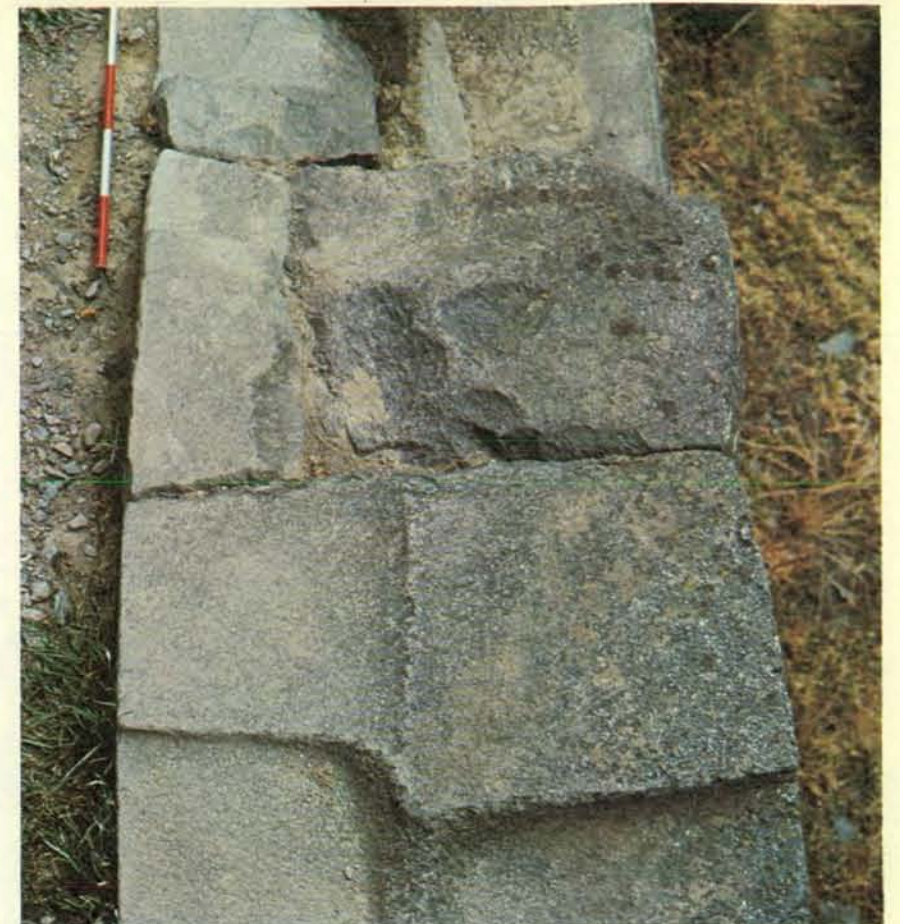
Le pietre usate come martelli e l'andesite delle pietre da costruzione hanno pressappoco la stessa durezza. La durezza si misura con la scala di Mohs. Nella scala di Mohs il talco, il minerale più tenero, ha durezza 1, mentre il diamante, il più duro, ha durezza 10. Le pietre da percussione da me trovate nella Cava dei lama hanno una durezza di 5,5 circa, pressappoco la stessa durezza dell'andesite dei blocchi da costruzione. Le pietre usate come martelli sono però più dure dell'andesite, la quale, avendo subito un raffreddamento differenziale durante la formazione ha accumulato al suo interno delle tensioni. Quando l'andesite viene colpita, le tensioni si liberano provocando la frammentazione della roccia. Di conseguenza, i ciottoli di fiume possono essere utilizzati efficacemente come martelli per squadrare e rifinire le pietre da costruzione.

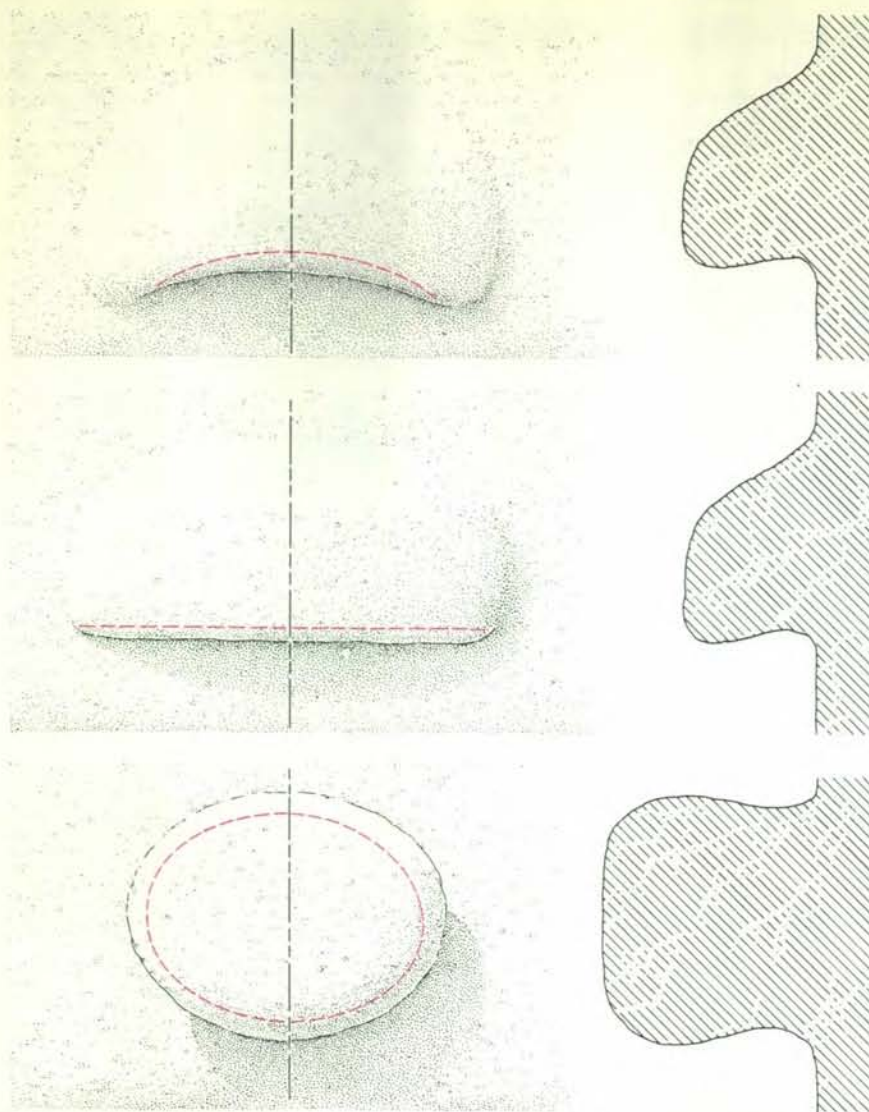
Evidentemente i tagliapietre inca usavano «martelli» di dimensioni diverse nelle varie fasi del processo di produzione dei conci. Nella ricerca che ho condotto nei siti delle cave ho trovato tre gruppi di percussori. Il primo gruppo includeva pietre con un peso compreso tra otto e 10 chilogrammi, il secondo pietre con un peso compreso tra due e cinque chilogrammi e il terzo pietre di peso inferiore a un chilogrammo. Sono convinto che ogni gruppo avesse una funzione specifica. I percussori più grossi potrebbero essere serviti per il lavoro di sgrossatura e squadratura dei blocchi dopo il loro distacco dalla superficie della roccia. La maggior parte dei blocchi non rifiniti presenta tracce di scheggia-

Un esperimento compiuto dall'autore rivela come i tagliapietre incaici avrebbero squadrato i blocchi da costruzione. Dopo aver grosso modo squadrato un blocco di andesite (una roccia vulcanica) prelevato dalla Cava dei lama, l'autore ha preso una pietra di circa quattro chilogrammi e l'ha usata per martellare una delle sei facce (*1*). Il percussore era tenuto senza eccessivo sforzo fra le mani, che si limitavano a guidarne la caduta sulla pietra a un angolo di 15-20 gradi dalla normale. Immediatamente prima di colpire la pietra, al percussore veniva impartita un'ulteriore inclinazione con il polso, cosicché l'angolo diveniva di circa 45 gradi (*2*). Dopo ogni colpo, il percussore rimbalzava di circa 25 centimetri (*3*). Terminata la lavorazione della prima faccia, il blocco era lasciato nella stessa posizione e si passava a un percussore più piccolo per rifinire gli spigoli della faccia successiva (*4*). Il percussore più piccolo, che pesava 560 grammi, era tenuto saldamente in mano, per impartire colpi di striscio dallo spigolo verso l'esterno. Poi il blocco era ruotato in modo da usare sulla seconda faccia il percussore più grosso (*5*). La tecnica ha dato come risultato un blocco dagli angoli leggermente convessi, molto simili a quelli che si osservano nei blocchi in pietra degli inca (*6*).

tura caratteristiche, simili a quelle che si osservano sugli utensili in pietra scheggiata, ma molto più grandi. I segni di scheggiatura sono stati probabilmente prodotti dai grossi percussori, usati per la squadratura. I percussori medi servivano probabilmente per rifinire le superfici dei blocchi e i più piccoli per gli spigoli.

Per accertare se i tagliapietre inca potessero avere usato i tre gruppi di percussori in questo modo sono passato dall'osservazione alla sperimentazione. La materia prima del mio esperimento è stato un pezzo grezzo di andesite che misurava circa 25 × 25 × 30 centimetri. Usando come percussore una pietra di quattro chilogrammi circa, ho eliminato le sporgenze maggiori per dare al blocco di andesite la forma approssimativa di un parallelepipedo. Per raggiungere lo scopo sono stati sufficienti sei colpi. L'obiettivo successivo è stato quello di spianare una delle sei facce del blocco. A questo scopo ho scelto un percussore diverso, sempre del peso di quattro chilogrammi circa, e con esso ho cominciato a martellare il blocco. Si potrebbe pensare che impugnare per un periodo prolungato una pietra di quattro chilogrammi sia faticoso. Il lavoro è facilitato, invece, dalla gravità. Senza bisogno di impugnare il percussore con forza, lo si può lasciar cadere sulla superficie del blocco, guidandolo semplicemente con entrambe le mani. Agendo in questo modo il percussore rimbalzerà sul blocco di andesite di 15-25 centimetri cosicché lo si potrà lasciar cadere di nuovo su di esso. L'operazione può così essere ripetuta per un lungo periodo e lo sforzo richiesto è piccolo.

Lavorando in questo modo si darebbe però forma al blocco essenzialmente rompendo la pietra. Se invece il percussore viene diretto sulla superficie del blocco a un angolo compreso fra 15 e 20 gradi dalla normale (perpendicolare), si staccano piccole schegge e l'operazione di taglio viene molto accelerata. Ho trovato che l'efficienza del colpo può essere ulteriormente aumentata dando al percussore una maggiore inclinazione con il polso subito prima che esso entri in contatto con la superficie del blocco. La torsione del polso fa aumentare l'angolo di impatto a 40-45 gradi rispetto alla normale (*si veda l'illustrazione nella pagina a fronte*). Il meccanismo con il quale aumentando l'angolo aumenta l'efficacia del taglio si può spiegare facilmente. Quando il percussore è calato verticalmente, la forza del colpo si converte in compressione, la quale frantuma la roccia. Se il colpo è deviato dalla verticale, alla compressione si aggiunge una forza di taglio, la quale aumenta all'aumentare dell'angolo del colpo: è la forza di taglio a staccare le minuscole schegge di pietra, accelerando l'operazione.

Dopo che una delle sei facce del blocco è stata spianata, il tagliapietre deve mutare tecnica. Se il blocco venisse semplicemente ruotato e si usasse lo stesso percussore per squadrare la nuova faccia, senza dubbio dagli spigoli di questa i colpi del grosso percussore staccherebbero grandi scheg-

Questo muro demolito a Ollantaytambo fornisce indizi sul modo in cui gli inca facevano combaciare i conci. Ogni depressione concava indica il punto da cui è stata tolta una pietra sovrastante. Le depressioni venivano eseguite con precisione martellando il blocco con una pietra sino a ottenere una corrispondenza perfetta con la superficie inferiore della pietra che vi veniva posata sopra.

I segni di trascinamento sulla faccia inferiore di un blocco trovato a Ollantaytambo suggeriscono che alcune pietre fossero trasportate fino al sito di costruzione sulle strade incaiche ricoperte di ghiaia. I segni possono essere analizzati per determinare la direzione in cui il blocco veniva trascinato. Per esempio, la depressione circolare (*a sinistra del centro*) è delimitata in modo netto a sinistra e in modo smussato a destra. Quando il blocco veniva trascinato, la ghiaia finiva nella depressione sotto al bordo anteriore, che rimaneva netto. Raggiunta la parte posteriore della depressione, essa era compressa fra fondo stradale e bordo posteriore della rientranza, che si usurava e smussava. Questo fa pensare che il blocco fotografato sia stato trascinato verso sinistra.

Le sporgenze sui blocchi da costruzione degli inca hanno varie forme, corrispondenti probabilmente a funzioni specifiche. Un tipo (*in alto*) è adatto all'applicazione di leve. Un altro (*in basso*) sarebbe servito a legarvi funi. Un terzo (*al centro*) potrebbe avere assolto entrambe le funzioni.

ge. Per evitare l'inconveniente, il tagliapietre deve passare a un percussore più piccolo per formare gli spigoli della nuova faccia prima che venga spianata la parte centrale. Per questo lavoro ho usato una pietra del peso di circa 560 grammi. Il metodo è del tutto diverso da quello del taglio della faccia: anziché colpire la superficie del blocco più o meno verticalmente, il percussore colpisce lo spigolo di striscio. La gravità ha poca parte nel lavoro compiuto sullo spigolo del blocco. Il percussore da 560 grammi è troppo piccolo per essere lasciato cadere e fatto poi rimbalzare. Esso deve essere tenuto saldamente in mano e la forza del colpo viene esclusivamente dal braccio del tagliapietre.

Una volta che siano stati formati gli spigoli, il blocco può essere ruotato. Il percussore piccolo viene messo da parte e il tagliapietre riprende il percussore più pesante per spianare la nuova faccia. Sul mio concio ho spianato due facce dopo la prima provando vari altri percussori di peso compreso fra 3,5 e quattro chilogrammi. A lavoro ultimato, ho ottenuto un blocco abbastanza ben squadrato. L'intero processo, dalla squadratura del blocco alla formazione di cinque spigoli e alla spianatura di tre facce, non ha richiesto più di 90 minuti. Il mio esperimento dimostra che le pietre da costruzione possono essere estratte, sgrossate e squadrate usando utensili semplici, in un modo che richiede poco tempo o sforzo. Il problema successivo è vedere se questi siano realmente i metodi usati dagli inca.

Le prove di ordine fisico in base alle quali gli inca avrebbero usato tecniche simili a quelle da me sperimentate sono numerose. Sulle pietre di tutti i muri incaici, a prescindere dal tipo di roccia, si trovano segni del distacco di schegge simili a quelli lasciati dal mio lavoro sul concio sperimentale. Se il blocco è di calcare, nel segno lasciato dal distacco della scheggia o attorno a esso vi sarebbe uno scoloramento biancastro. Le chiazze bianche indicano senza dubbio una parziale metamorfosi del calcare conseguente al calore generato dalla percussione. Su ogni pietra da me esaminata i segni di scheggiatura sono più piccoli verso gli spigoli che al centro delle facce, a dimostrazione del fatto che i percussori usati per formare gli spigoli erano più piccoli di quelli usati al centro delle facce. Altri dati provengono dal cronista coevo Garcilaso de la Vega, noto come «*el Inca*». Figlio di un *conquistador* e di una principessa inca, egli scrisse nel 1609 che gli inca «non hanno per lavorare la pietra altri utensili se non alcune pietre nere... con le quali squadrano i blocchi martellandoli piuttosto che tagliandoli».

La questione forse più interessante di tutte riguarda non l'estrazione o la squadratura dei blocchi, ma il modo in cui le grandi pietre venivano fatte combaciare con tanta precisione. Le commessure sono di due tipi principali: di base e laterali. Quelle di base sono le giunzioni orizzontali attraverso le quali la maggior parte del peso di un blocco viene trasmesso al corso, o serie di pietre, sottostante. Le commessure laterali sono le giunzioni fra pietre dello stesso corso e attraverso di esse si trasmette poco peso o niente affatto. Mi occuperò qui in particolare delle giunzioni di base.

Dopo avere esaminato molti muri incaici sono giunto alla conclusione che, quando ne veniva costruito uno, le commessure di base venivano realizzate lavorando la faccia superiore dei blocchi del corso già messo in opera, sul quale venivano appoggiate a secco le pietre del corso successivo. Le superfici delle pietre erano in generale lievemente convesse e la faccia superiore delle pietre già in opera, che dovevano accoglierle, veniva perciò lavorata in modo da renderla concava. Dovunque ci si imbatta in un muro demolito, si possono osservare chiaramente le depressioni concave nelle pietre dei corsi rimasti in piedi; in esse si rilevano precise impronte lasciate dalla parte inferiore delle pietre rimosse (*si veda l'illustrazione in alto a pagina 63*). Queste depressioni concave non concordano con un'ipotesi che è stata spesso suggerita riguardo alle opere murarie incaiche, ossia che le pietre contigue venissero sfregate l'una contro l'altra per ottenere un perfetto combaciamento. È chiaro che, sfregando due superfici fra loro non è possibile conseguire giunzioni perfette concavo-convesse come quelle da me osservate. In quale modo dunque veniva ottenuta questa mirabile corrispondenza?

Come per la squadratura delle pietre, ho tentato di cimentarmi nel compito di imparare in quale modo lavorassero gli inca per riuscire a fare combaciare perfettamente i conci. In questo esperimento ho utilizzato il blocco di andesite già usato nell'esperimento di squadratura e un bloc-

co più grosso, in cui scavare la concavità della commessura di base che doveva accogliere la pietra superiore. Cominciai il lavoro ponendo il blocco più piccolo su quello più grande e disegnandone il contorno. Tolto il blocco più piccolo e, usando il contorno come guida, ho scavato nel blocco più grande, con un lavoro di martellatura, una depressione corrispondente alla forma complessiva della faccia inferiore della pietra più piccola. Il lavoro di percussione produceva abbondante polvere, che doveva essere spazzata via. La polvere ostacola, infatti, il lavoro perché smorza i colpi del percussore; tuttavia essa è anche utile: quando il blocco superiore viene posato di nuovo per prova sul blocco sottostante, lascia nella polvere un'impronta della sua superficie inferiore. La polvere viene compressa nei punti di contatto, mentre non è compressa affatto dove le due pietre non combaciano. Tolta la pietra, si continua a colpire nei punti di contatto, indicati dalle aree dove la polvere è stata compressa. Ripetendo varie volte questo procedimento, è possibile ottenere alla fine una corrispondenza tanto precisa quanto la si desidera.

La stessa tecnica può essere utilizzata per formare le commessure laterali. Il nuovo concio che deve essere aggiunto al corso viene accostato ai blocchi già in opera e su questi si scavano depressioni concave. Le commessure laterali differiscono da quelle di base per il fatto che la corrispondenza perfetta che si osserva guardando il muro frontalmente prosegue, a volte, per pochi centimetri, mentre la parte più interna della commessura è riempita da pietrisco. In molti casi, però, le commessure laterali combaciano sull'intera superficie di contatto con la stessa perfezione delle commessure di base.

È chiaro che la tecnica usata dagli inca per far combaciare i blocchi di pietra consisteva in gran parte nel procedere per tentativi ed errori. Questo è un metodo laborioso, specialmente se si considerano le dimensioni colossali di alcuni dei blocchi di pietra usati a Sacsahuamán e a Ollantaytambo. Si deve però tener presente che tempo e mano d'opera non costituivano un problema per gli inca, i quali non avevano una nozione europea del tempo e disponevano di una grande abbondanza di mano d'opera che era il tributo dei popoli conquistati. I miei esperimenti dimostrano, inoltre, che con un poco di pratica si fa presto l'occhio alle corrispondenze fra superfici, cosicché il tempo necessario per ottenere commessure soddisfacenti si riduce di molto. A favore del mio metodo si può sottolineare che esso funziona e, inoltre, non richiede utensili diversi da quelli effettivamente documentati. Esso ha in più il sostegno di almeno due autori del Cinquecento. Uno di loro, José de Acosta, un gesuita che viaggiò con i *conquistadores* spagnoli ed è considerato un osservatore molto attendibile, scrisse nel 1589: «Tutto questo fu realizzato con molta mano d'opera e molta pazienza nel lavoro, poiché per adattare una pietra all'altra, finché combaciassero perfettamente, era necessario compiere molte prove.»

Io ritengo che i miei esperimenti forniscano una spiegazione ragionevole del metodo con cui i tagliapietre inca estraevano le pietre, le squadravano e le facevano combaciare. In che modo i blocchi venissero trasportati al sito di costruzione e come venissero manovrati, una volta giunti alla sede definitiva, sono interrogativi che non hanno invece trovato finora alcuna risposta soddisfacente.

Nel manovrare i massi, che erano spesso molto pesanti, avevano senza dubbio una funzione importante le sporgenze lasciate sulle loro superfici. Con una grande varietà di dimensioni e di forme (*si veda l'illustrazione alla pagina 64*), queste protuberanze si trovano in generale sulla parte inferiore di un blocco già messo in opera e potrebbero essere state utilizzate per attaccarvi delle corde, o come punti ai quali applicare la forza di una leva. A quanto pare, esse venivano scolpite sui conci soltanto nel sito di costruzione e la loro funzione specifica doveva essere quella di facilitare la manovra dei massi. Poiché nessuno dei blocchi di pietra abbandonati lungo le strade di trasporto presenta sporgenze, sembrerebbe che queste non avessero alcuna funzione nel trasporto dei blocchi al sito di costruzione.

Come venivano trasportati i blocchi? Qualche indizio preliminare è fornito dai blocchi sparsi, che si trovano a Ollantaytambo. Su questi blocchi si può osservare un particolare logorio contrassegnato da striature longitudinali più o meno parallele. Tanto il logorio quanto le striature sembrano essere il risultato del trascinamento dei blocchi dalle cave al sito di costruzione. La direzione in cui il blocco veniva trascinato può essere determinata facilmente dai segni. Se si esamina attentamente la superficie di un blocco, si scoprono aree dalla forma irregolare che non sono state usurate perché sono leggermente rientranti. Queste aree hanno in generale una delimitazione netta su un lato e una delimitazione smussata, graduale, sull'altro. Quando il masso veniva trascinato, il bordo netto si trovava nella parte anteriore e quello smussato nella parte posteriore. La ghiaia del fondo stradale doveva accumularsi nella parte posteriore della depressione e doveva rimanere compressa fra il blocco e la strada, logorando in tal modo l'area del bordo posteriore.

Altre caratteristiche dei blocchi contribuiscono a darci un quadro più completo di come essi venissero trasportati. Il logorio si trova solo sulla superficie più grande della pietra, suggerendo che i blocchi venissero trascinati nella loro posizione più stabile. I blocchi trovati nella cava non presentano logorio e l'estensione della superficie usurata aumenta con la distanza dalla cava. La presenza del logorio toglie verosimiglianza all'ipotesi che gli inca trasportassero le pietre più grandi su rulli o scivoli. Le tracce di trascinamento non escludono la possibilità che rulli o scivoli venissero usati nelle parti delle rampe situate più in alto sulla collina, ma non è stata trovata alcuna traccia materiale di tali attrezzi.

Se i blocchi venivano trascinati lungo le rampe di accesso, gli inca devono aver dedicato a questo compito una mano d'opera considerevole, particolarmente per i blocchi più grossi. La forza richiesta per trascinare un blocco dipende dal coefficiente d'attrito tra la pietra e il materiale della rampa, dalla pendenza della rampa e dal peso del blocco. Ho determinato sperimentalmente il coefficiente d'attrito e ho trovato che la pendenza della rampa di Ollantaytambo era di circa 10 gradi. A Ollantaytambo il blocco più grosso pesa circa 140 tonnellate. Ho calcolato che per trascinare un blocco del genere su per la rampa occorrerebbe una forza di circa 120 400 chilogrammi. Ammesso che un uomo possa esercitare una trazione costante con una forza di 50 chilogrammi (che potrebbe essere un stima sbagliata per eccesso), occorrerebbero circa 2400 uomini per trascinare il blocco sino alla cima della rampa. Questa cifra concorda (almeno come ordine di grandezza) con la testimonianza dell'autore cinquecentesco Pedro Cieza de Léon, il quale osservò che, dei 20 000 uomini assegnati alla costruzione di Sacsahuamán, 6000 erano incaricati del trasporto.

Quanto si è detto fin qui sembra ragionevole, ma non sgombra il terreno da interrogativi ai quali non sono stato in grado finora di dare una risposta. Le rampe degli inca erano larghe solo da sei a otto metri e io non sono riuscito a trovare soluzioni plausibili a due problemi posti da questa dimensione ridotta. Uno è in che modo 2000 o più uomini potessero essere «imbrigliati» in modo che ciascuno di loro potesse cooperare al trascinamento del blocco. L'altro è in che modo una folla così grande potesse trovare posto su una rampa tanto stretta. Questi sono solo due dei problemi irrisolti concernenti il trasporto dei blocchi. Fra gli altri vi sono le tecniche per legare le funi ai blocchi e i metodi per manovrare i colossali blocchi di pietra.

Inoltre, le pietre provenienti da Rumiqolqa probabilmente non venivano trascinate. Diversamente dai blocchi di Kachiqhata, quelli di Rumiqolqa venivano squadrati con cura prima di lasciare la cava. Su di essi non si trova alcuna traccia di trascinamento e pare del resto irragionevole pensare che una faccia ben rifinita venisse trascinata su una rampa di pietre. Come venivano dunque trasportati i blocchi perfettamente squadrati? Prima che si possa dare una versione definitiva alle tecniche di costruzione in pietra degli inca si dovrà trovare una risposta a questi e a molti altri interrogativi. Quindi gli studi sono tutt'altro che terminati. Per ora, la sperimentazione e l'osservazione ci hanno permesso di dare molte risposte ad alcuni fra i quesiti fondamentali concernenti l'estrazione, la squadratura e il perfetto combaciamento delle pietre da costruzione.

Il cuore come ghiandola endocrina

Oltre che funzionare come pompa, questo importante organo corporeo produce un ormone scoperto recentemente, il fattore natriuretico atriale, che interagendo con altri ormoni regola la pressione e il volume sanguigni

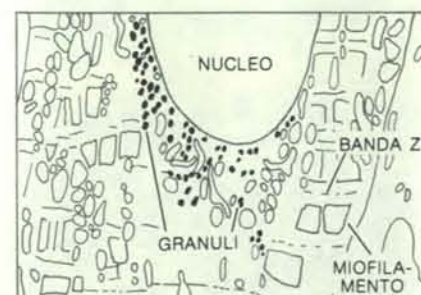
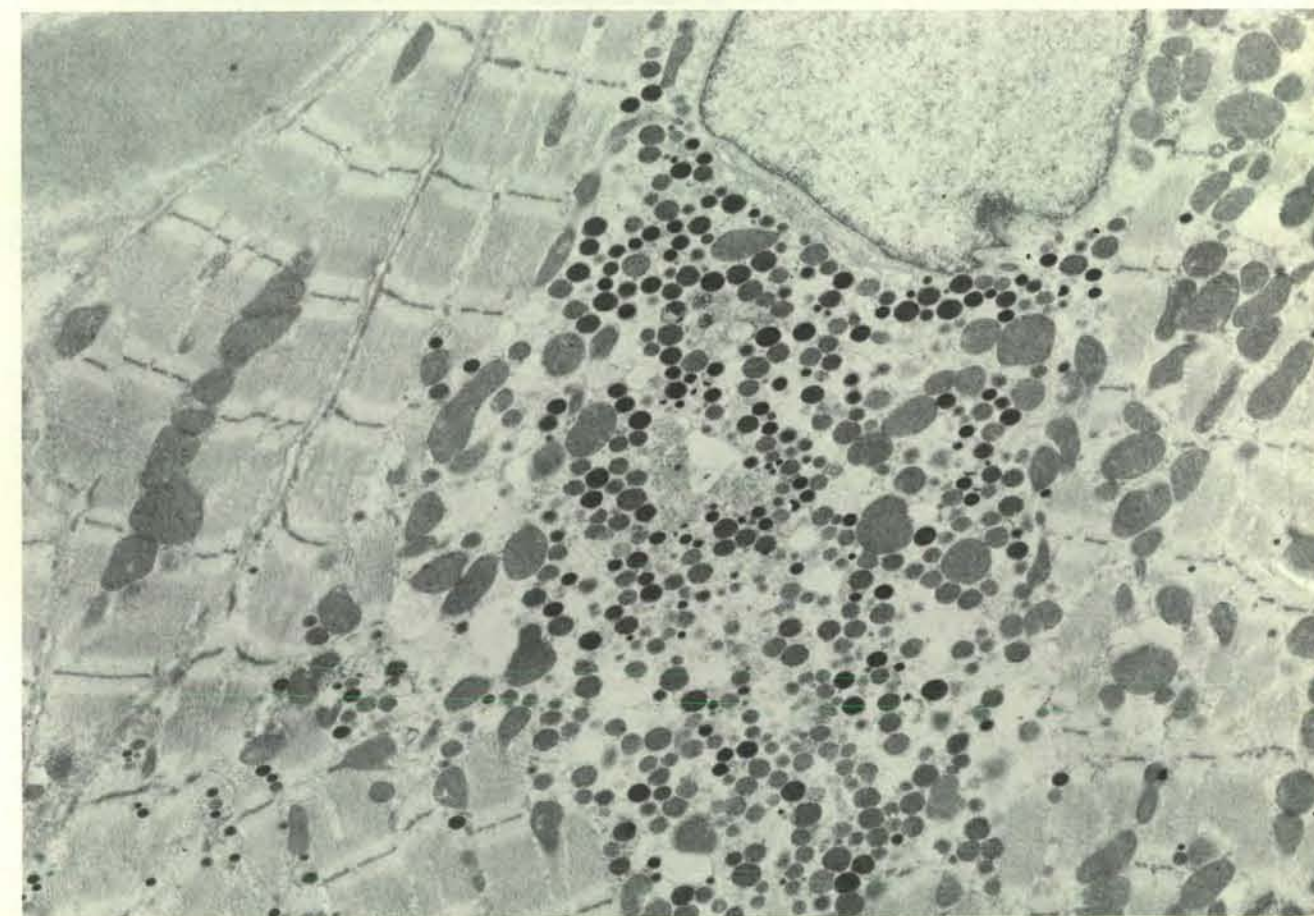
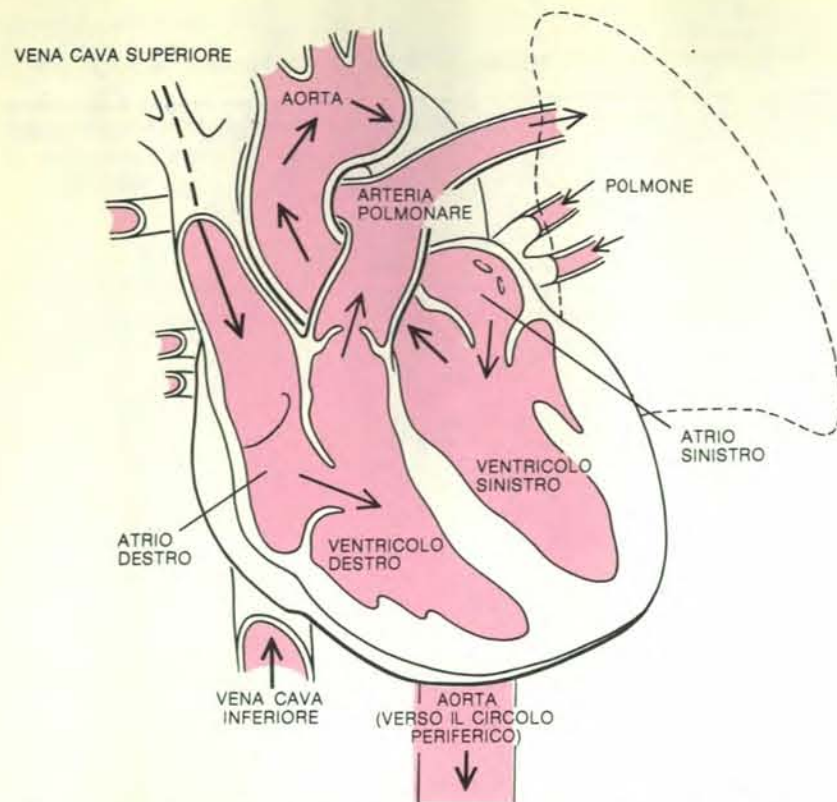
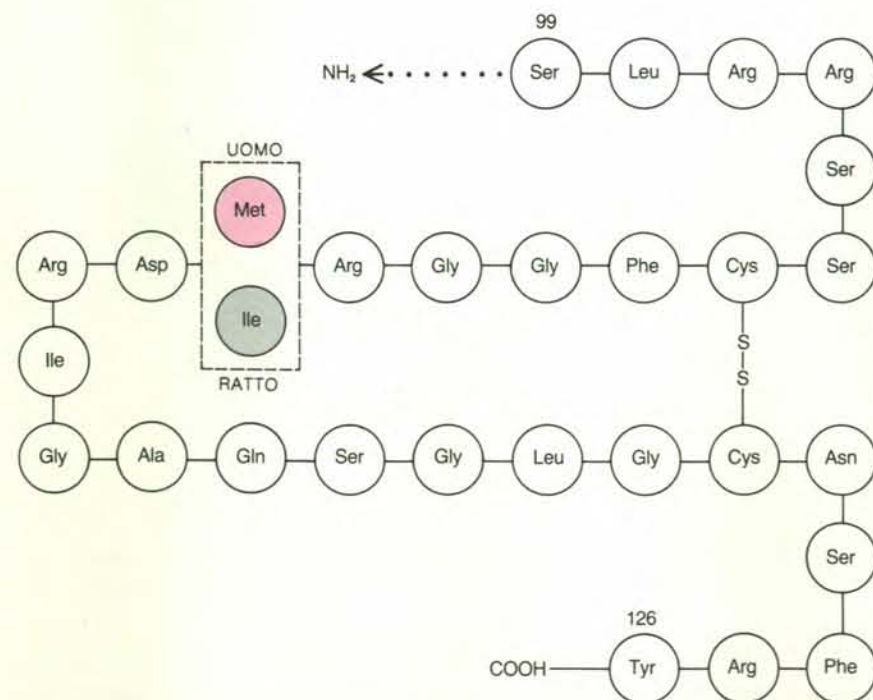
di Marc Cantin e Jacques Genest

Il cuore è una pompa, un organo muscolare che si contrae ritmicamente, spingendo il sangue prima verso i polmoni per l'ossigenazione e poi verso il sistema vascolare per rifornire di ossigeno e di sostanze nutritive ogni cellula dell'organismo. Tutto ciò è noto fin da quando, nel 1628, William Harvey pubblicò l'*Essay on the Motion of the Heart and of the Blood in Animals.*

Da qualche anno, però, è noto che il cuore è qualcosa di più di una pompa: è anche una ghiandola endocrina. Infatti, esso secerne un potente ormone peptidico, il fattore natriuretico atriale (ANF dall'inglese *atrial natriuretic factor*). Questo ormone ha un ruolo importante nella regolazione della pressione e del volume ematici nonché dell'escrezione di acqua, sodio e potassio e agisce su svariati organi: sugli stessi vasi del sangue, sui reni e sulle ghiandole surrenali, oltre che su aree del cervello che hanno funzione regolatrice.

La recente scoperta dell'ANF ha risolto un problema che esisteva da molto tempo. Infatti, fin dal 1935 John Peters della School of Medicine della Yale University aveva intuito che doveva esserci un meccanismo situato all'interno del cuore o nelle sue vicinanze e capace di «misurare il grado di riempimento dei vasi» e di regolare in modo accurato il volume sanguigno. Negli anni cinquanta e sessanta molti ricercatori avevano cercato inutilmente un ipotetico «ormone natriuretico». La sua presenza avrebbe potuto spiegare la natriuresi (escrezione di sodio) e la contemporanea diuresi (escrezione di acqua), riscontrate in assenza di variazioni a carico dei noti processi che le regolano. Questo tipo di natriuresi e diuresi non spiegabili si potevano osservare in seguito a distensione degli atri, le due cavità superiori del cuore, che ricevono il sangue dalle vene polmonari o dalle vene cave e lo trasferiscono ai vicini ventricoli. L'ormone era stato chiamato «terzo fattore», in quanto avrebbe completato il ruolo degli altri due noti meccanismi regolatori della pressione e del volume sanguigni, e cioè l'attività dell'aldosterone, un ormone, e la filtrazione del sangue da parte del rene.

Il primo passo verso la scoperta del terzo fattore è stato compiuto nel 1956, quando Bruno Kisch dell'American College of Cardiology ha notato la presenza di «corpi densi», come egli li ha definiti, all'interno dei cardiociti, o cellule muscolari cardiache, di atri di cavie. Nel 1964, James D. Jamieson e George E. Palade della Yale School of Medicine hanno pubblicato che questi corpi, la cui funzione era ancora sconosciuta, erano apparentemente presenti negli atri di tutti i mammiferi presi in esame, inclusa la specie umana. Nel 1974, il nostro gruppo che lavora all'Università di Montreal ha notato che essi erano molto simili ai granuli di secreto accumulati nelle cellule endocrine (cioè le cellule che secernono ormoni), per esempio le cellule del pancreas o del lobo anteriore dell'ipofisi, e ha osservato, inoltre, che quando si somministravano agli animali amminoacidi marcati con isotopi radioattivi, in breve questi comparivano nei granuli degli atri, incorporati in polipeptidi appena sintetizzati. Questo è esattamente ciò che sarebbe accaduto nel caso dei granuli di secreto di cellule endocrine.

Nel 1976 Pierre-Yves Hatt e collaboratori dell'Università di Parigi hanno messo in relazione le conoscenze acquisite fino a quel momento sui granuli con le osservazioni precedenti, che riguardavano la regolazione dei livelli di sodio e di acqua. Essi sono così riusciti a dimostrare che il numero di granuli dei cardiociti atriali aumenta parallelamente alla riduzione della quantità di sodio nella dieta dell'animale, il che significa che i granuli dovevano accumulare qualche sostanza correlata con la regolazione dell'equilibrio del sodio. Un grande passo avanti è stato compiuto quando, nel 1981, Adolfo J. de Bold, Harald Sonnenberg e collaboratori alla Queen's University di Kingston nell'Ontario hanno iniettato omogeneizzati di atri di

Nelle fotografie al microscopio elettronico, effettuate nel laboratorio dell'autore, appaiono, ingranditi 12 000 volte, granuli di accumulo di ormone in cellule muscolari cardiache (cardiociti) di ratto. La scoperta dei granuli ha portato all'ipotesi che il cuore sia una ghiandola endocrina. Lo stiramento dell'apparato contrattile di un cardiocito (*filamento con bande* Z) stimola la liberazione del fattore natriuretico atriale (ANF). In una cellula di ratto normale (*in alto*) i granuli sono tutti raggruppati vicino al nucleo. In una cellula di ratto, alimentato per 30 giorni con una dieta carente di sodio (*in basso*), il numero dei granuli è aumentato, probabilmente per la diminuzione del volume del sangue. Infatti, ciò determina l'abbassamento del livello dell'ANF in circolo; a questo evento segue poi un accumulo di granuli nella cellula.

Nei mammiferi l'ANF è secreto dalle cellule muscolari dei due atri del cuore. Il sangue povero di ossigeno, proveniente dalla periferia, entra nell'atrio destro, portato dalle vene cave; da qui si riversa nel ventricolo destro ed è quindi spinto verso i polmoni attraverso l'arteria polmonare. Il sangue ricco di ossigeno ritorna all'atrio sinistro, passa nel ventricolo sinistro ed è spinto nell'aorta, da cui viene poi distribuito alla periferia. In alcune specie diverse dai mammiferi sembra che anche i ventricoli abbiano attività diuretica e natriuretica; può darsi quindi che anch'essi secernano l'ANF.

La sequenza di amminoacidi nella molecola di ANF in circolo sembra identica nella specie umana e nel ratto, con la sola eccezione dell'amminoacido in posizione 110, che è metionina (Met) nell'ANF umano e isoleucina (Ile) nell'ANF di ratto. Le molecole di ANF attivo sia umano sia di ratto consistono di 28 amminoacidi con un legame disolfuro tra due residui di cisteina (Cys), indispensabile per l'attività dell'ANF. L'ormone attivo che si trova in circolo proviene dalla scissione di un polipeptide precursore più lungo, formato da 152 amminoacidi nel ratto e da 151 nell'uomo.

ratto in altri ratti e hanno notato la successiva rapida comparsa di diuresi e natriuresi massicce e di breve durata. Essi hanno concluso quindi che gli atri dovevano effettivamente contenere un «fattore» in grado di promuovere questi effetti e gli hanno dato il nome di fattore natriuretico atriale.

Durante i tre anni successivi l'ANF è stato localizzato per la prima volta con sicurezza e identificato a livello biochimico. Si è notato che, nei ratti, i granuli sono da due a due volte e mezzo più abbondanti nell'atrio destro che nell'atrio sinistro e sono fortemente concentrati vicino alla superficie del cuore e nelle zone più esterne degli atri. Essi non sono stati ritrovati nei ventricoli di ratto o di altri mammiferi e l'iniezione di estratti di ventricolo di mammifero non ha alcun effetto sui vasi sanguigni, sulla diuresi e sulla natriuresi. In specie diverse dai mammiferi, i granuli sono stati invece ritrovati sia nei ventricoli sia negli atri e, anche in queste specie, sono da mettersi in relazione con effetti diuretici e natriuretici. La presenza dei granuli nei ventricoli di specie diverse dai mammiferi è in accordo con il fatto che le cellule del cuore tendono a essere più specializzate nelle specie superiori.

L'ANF, dopo la sua localizzazione, avvenuta nel giugno del 1983, è stato isolato e purificato dal nostro gruppo. Due mesi dopo è stato sintetizzato da Ruth F. Nutt e collaboratori dei Merck Sharp & Dohme Research Laboratories. L'ANF è la porzione attiva di una molecola più grande che ne è il precursore. Tutti i diversi gruppi di ricercatori che hanno determinato la sequenza di amminoacidi del polipeptide hanno trovato che l'ANF aveva sempre lo stesso corpo centrale di 21 amminoacidi. La forma attiva dell'ormone, che si trova in circolo nel ratto, ha complessivamente 28 amminoacidi e un peso molecolare di 3060. La parte attiva dell'ormone è unita a un peptide inattivo di 100 amminoacidi e a un peptide di segnale, di 24 amminoacidi, che viene staccato quando la molecola viene sintetizzata. La forma dell'ormone che si trova in circolo nella specie umana non è stata ancora determinata, anche se noi riteniamo che anche in questo caso consti di 28 amminoacidi.

Recentemente il gene che codifica per l'ANF nella specie umana è stato clonato e ne è stata determinata la sequenza, in modo da poter sintetizzare l'ormone o per via chimica o introducendo il gene in lieviti o batteri; con entrambi i metodi si può produrre una quantità di ormone sufficiente per consentire lo studio dei suoi effetti sull'intero organismo. Sono stati prodotti, inoltre, degli anticorpi diretti contro l'ormone: questo consente oggi di utilizzare alcuni dosaggi immunologici molto sensibili e che permettono di individuare la secrezione dell'ANF nonché i siti in cui esso fa sentire il suo effetto.

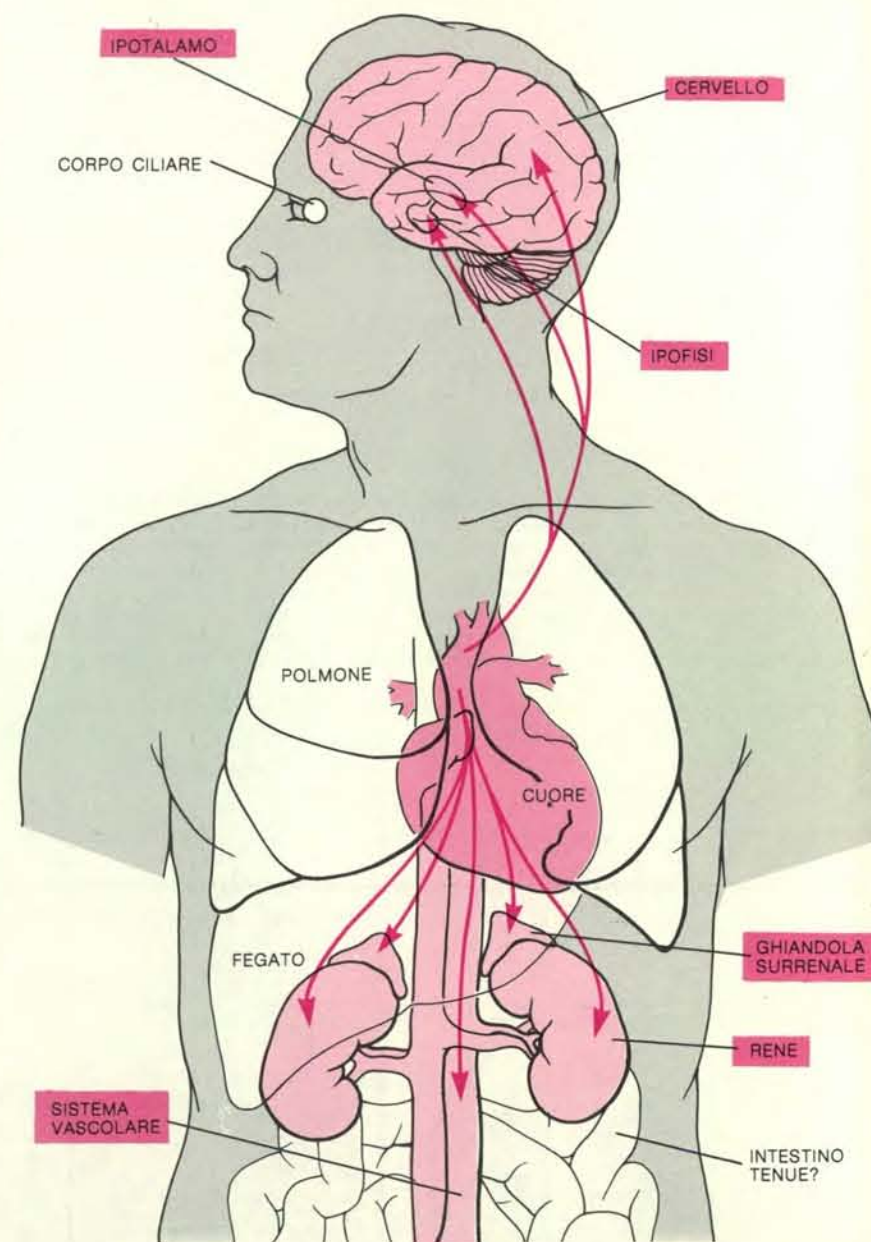
Quando i ratti vengono costretti all'immobilità, nel loro sangue l'ANF aumenta da cinque fino a 20 volte. La liberazione dell'ANF da parte del cuore è stata anche misurata in pazienti affetti da malattie valvolari e con un volume ematico superiore alla norma; questi pazienti erano sottoposti a cateterismo cardiaco, una tecnica che permette di prelevare un campione di sangue dalle arterie e dalle cavità cardiache. È stato notato che la concentrazione plasmatica di ANF era da due a otto volte più elevata nel sangue venoso del seno coronarico, che drena gli atri, che nel sangue arterioso o venoso periferico. Queste osservazioni confermano che l'ANF viene secreto principalmente, se non esclusivamente, dalle cellule degli atri. Sia nei ratti sottoposti allo stress dell'immobilità, sia nei pazienti cateterizzati, i cardiociti atriali sono soggetti a stiramento, ed è questa condizione il segnale per la liberazione di ANF. Anche un aumento del volume ematico può causare lo stiramento dei cardiociti atriali e la liberazione di ANF, come è stato dimostrato nel caso in cui il volume ematico era stato aumentato artificialmente con l'infusione di una soluzione salina.

Dopo che, in risposta allo stimolo di stiramento, i cardiociti hanno liberato l'ANF, questo peptide viene trasportato tramite il circolo arterioso agli organi bersaglio, e cioè ai reni, alle ghiandole surrenali, al cervello e a molti altri tessuti. In linea generale, l'effetto dell'ANF consiste nel modificare l'attività di un complicato circuito di retroazione omeostatico che regola la pressione sanguigna, il volume sanguigno e la ritenzione del sodio. Questo circuito è il sistema renina-angiotensina che collega tra loro alcune funzioni di cervello, cuore, arterie, ghiandole surrenali, reni e altri organi. Una delle sue sostanze chiave è la renina. La renina è un enzima secreto da cellule situate nelle arterie afferenti ai glomeruli e che formano strutture sacculari nel punto di ingresso nel rene. Queste cellule iuxtaglomerulari liberano la renina ogni volta che il livello del sodio nei tubuli renali distali è basso o è bassa la pressione locale del rene.

Una volta in circolo, la renina agisce sull'angiotensinogeno (un polipeptide) tagliandone via una porzione e trasformandolo così in angiotensina I, che è poi a sua volta trasformata in angiotensina II. Quest'ultima è un piccolo polipeptide che ha una potente azione costrittrice sulla muscolatura liscia dei vasi e un effetto di retroazione, in quanto sopprime parzialmente la secrezione di renina da parte delle cellule iuxtaglomerulari. Infine stimola le ghiandole surrenali a secernere aldosterone, un ormone che viene trasportato al rene e al lobo posteriore dell'ipofisi, facendo diminuire l'escrezione di sodio e acqua.

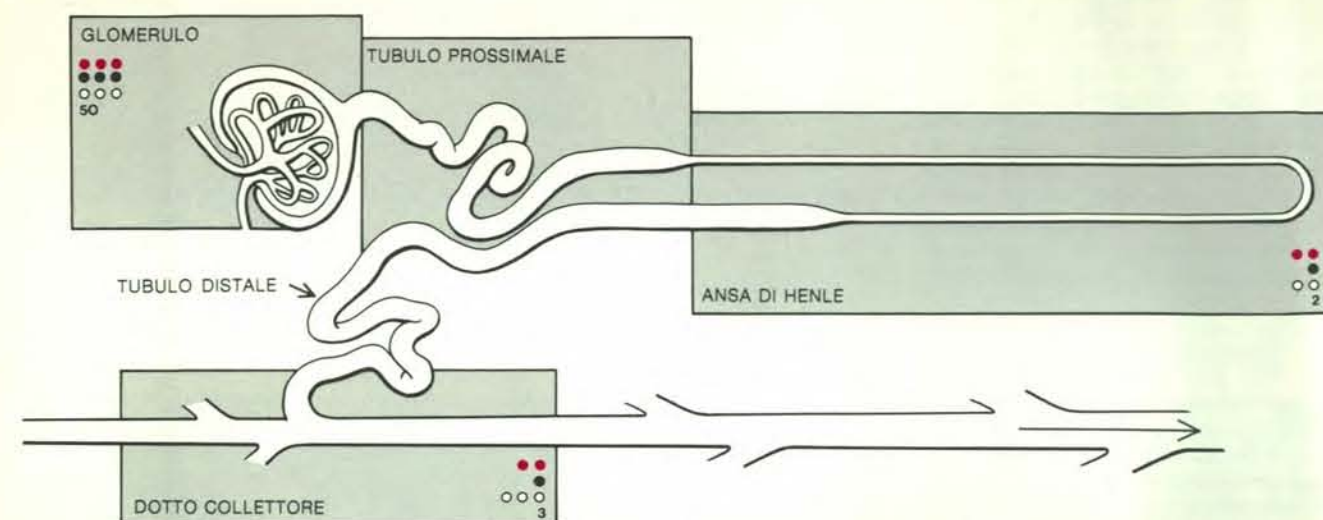
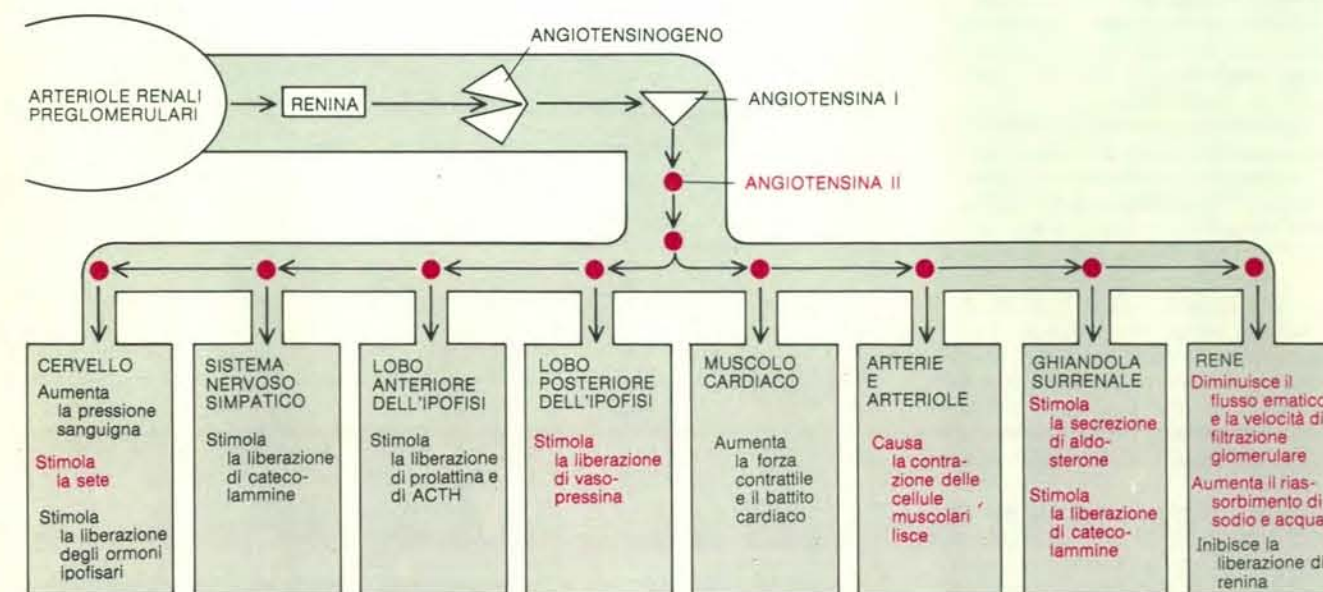
L'ANF influenza il sistema renina-angiotensina, inibendo in qualche modo la secrezione di renina e anche, direttamente, la secrezione di aldosterone da parte delle ghiandole surrenali. Il rapporto tra ANF e aldosterone è stato chiarito nel nostro laboratorio mediante una serie di esperimenti con cellule surrenali di bovino e di ratto in coltura. L'ANF inibiva del 20 per cento la produzione normale di aldosterone surrenalico e, inoltre, riduceva del 40-70 per cento l'aumento della produzione di aldosterone, che fa normalmente seguito alla stimolazione delle cellule surrenaliche con angiotensina II o con l'ormone ipofisario ACTH. Una riduzione analoga dei livelli di aldosterone plasmatico è stata notata in ratti e cani dopo iniezioni di ANF.

Per verificare se la diminuzione dei livelli di aldosterone era causata dalla presenza di siti di legame, specifici per l'ANF, sulla superficie delle cellule surrenali, abbiamo aggiunto dell'ANF marcato con isotopi radioattivi a colture di queste cellule, facendolo seguire poi da ANF «freddo», ovvero non marcato.

Abbiamo così notato che la concentrazione dell'ANF marcato diminuiva in maniera significativa dopo l'aggiunta dell'ANF freddo, segno questo che l'ANF freddo stava spiazzando il peptide marcato da molti siti di legame. Con l'uso di ACTH, angiotensina II e altri peptidi attivi, abbiamo poi scoperto che l'ANF si lega a siti cellulari molto specifici. Il fatto che questi

L'ANF agisce su varie regioni cerebrali (incluso l'ipotalamo), sul lobo posteriore dell'ipofisi, sulla ghiandola surrenale, sul rene e sul sistema vascolare. Inoltre si lega a bersagli nel polmone, nel fegato, nel corpo ciliare (che secerne l'umore acqueo dell'occhio) e, probabilmente, nell'intestino tenue, ma i suoi effetti su questi ultimi tessuti non sono chiari. Nel cervello l'ANF si lega a molte aree interessate al controllo della pressione sanguigna e alla regolazione del sodio e dell'acqua; nell'ipotalamo inibisce la liberazione di vasopressina, un ormone che si accumula nel lobo posteriore dell'ipofisi, è antidiuretico e ha effetto vasocostrittore su arteriole e capillari. L'ANF fa rilassare le cellule muscolari lisce dei vasi sanguigni, inibisce la secrezione di aldosterone (un ormone in grado di far aumentare la pressione sanguigna) da parte della ghiandola surrenale e stimola il rene ad aumentare l'escrezione del sodio e dell'acqua filtrati dal sangue attraverso i glomeruli. Inoltre, sembra faccia diminuire il passaggio del sodio dai tubuli distali e dai collettori renali al sangue.

peptidi non fossero in grado di esercitare lo stesso effetto spiazzante esercitato dall'ANF freddo dimostrava la presenza di siti di legame specifici per l'ANF.

Oltre agli effetti sul sistema renina-angiotensina, l'ANF ne provoca altri sul rene, regolando l'escrezione dell'acqua e del sodio. Campioni marcati con isotopi radioattivi, iniettati nell'aorta di ratto, dimostrano che l'ANF si lega alle cellule epiteliali dei glomeruli e a numerosi siti recettori nei vasi in prossimità dei glomeruli e dei tubuli. Inoltre, esercita in qualche modo un effetto di breve durata sul meccanismo con cui i glomeruli filtrano il sangue. È probabile che l'ANF faccia aumentare la permeabilità della parete dei capillari glomerulari in modo che maggiori quantità di acqua e di sodio siano filtrate dal sangue.

L'ANF agisce anche direttamente sui tubuli renali, dove si forma l'urina a partire dal plasma sanguigno dopo che questo è stato filtrato attraverso i glomeruli. I tubuli distali «riassorbono» il sodio dal filtrato e lo ritrasferiscono al sangue. È probabile che l'ANF diminuisca questa attività. Il meccanismo con cui esso agisce sui tubuli è tuttora sconosciuto, dal momento che non si è potuto dimostrare che richieda

Nel rene l'ANF sembra essere più attivo nei glomeruli e meno in altre regioni funzionali. Esso si lega a una cellula bersaglio, dove spesso attiva la guanilicociclasi particolata, un enzima localizzato sulla membrana cellulare e che attiva a sua volta il nucleotide guanosinmonofosfato ciclico (GMP ciclico), un secondo messaggero, che trasporta il messaggio dell'ANF a un sito all'interno della cellula. L'ANF inibisce anche la stimolazione dell'adenilicociclasi, il secondo messaggero di un altro sistema. Una serie di studi condotti usando ANF marcato con isotopi radioattivi hanno dimostrato un'intensità di legame molto elevata nei glomeruli di ratto (*pallini in colore*). Sulla stessa scala da 0 a 3 sono riportate altre prove, che hanno dimostrato un'accentuata attività guanilicociclasica (*pallini in nero*) e un'inibizione dell'attività adenilicociclasica (*pallini in bianco*) nei glomeruli. Il livello di GMP ciclico (*indicato da numeri*) è aumentato di oltre 50 volte rispetto al livello basale. I dotti collettori e la parte spessa delle anse di Henle hanno mostrato una certa attività legata all'ANF, mentre i tubuli prossimali non ne hanno mostrata affatto.

L'ANF inibisce alcuni effetti dell'angiotensina II, un ormone peptidico che, tra l'altro, fa aumentare la pressione e il volume del sangue. Quando la pressione del sangue nel rene è bassa, il rene riversa in circolo la renina. Questo enzima agisce su un polipeptide precursore, l'angiotensinogeno, staccandone una parte e producendo angiotensina I. Questa è a sua volta trasformata nell'angiotensina II, un peptide che il sangue porta ai diversi organi qui indicati. L'angiotensina II esercita i suoi effetti ipertensivi stimolando direttamente la contrazione dei vasi, inducendo la secrezione di altri ormoni che stimolano la contrattilità (come la vasopressina o l'adrenalina e le altre catecolammine) e sollecitando la ghiandola surrenale a secernere aldosterone, che agisce a livello renale e promuove la ritenzione di sale e di acqua. L'angiotensina II ha anche un effetto di retroazione, in quanto inibisce la liberazione della renina da parte del rene. Gli effetti che sono inibiti dall'ANF sono indicati in colore.

energia. Infatti, a differenza di altri meccanismi di riassorbimento, l'attività dell'ANF non comporta alcun consumo di ossigeno o demolizione di glucosio.

Il ruolo dell'ANF nel rilassamento delle cellule muscolari in tutto il sistema vascolare è altrettanto importante dell'effetto che l'ANF esercita sui glomeruli e sui tubuli renali. Sebbene l'iniezione di ANF sia in grado di rilassare e di far dilatare grosse arterie, quali le arterie vertebrali, le arterie femorali, le carotidi comuni e le coronarie, l'effetto più marcato del peptide si osserva nelle arteriole renali. In esperimenti in cui ANF sintetico veniva aggiunto al mezzo di incubazione contenente frammenti di arterie renali di ratto e di coniglio, si è avuta una marcata inibizione, per un periodo da 30 a 80 minuti, del classico effetto costrittore dell'angiotensina II e dell'adrenalina sul tessuto vascolare. Anche l'infusione di ANF in un rene isolato di ratto ha determinato una rapida caduta della pressione di perfusione della durata di 18 minuti. Ciò indica che l'ANF agisce o sulle cellule muscolari lisce dei vasi sanguigni o sull'endotelio, lo strato cellulare che delimita internamente i vasi. Quando il peptide è stato aggiunto ai vasi dopo che l'endotelio era stato distrutto, l'effetto di rilassamento vasale persisteva, dimostrando quindi che l'ANF agisce sulle cellule muscolari lisce.

Il meccanismo di azione dell'ANF sulle cellule muscolari lisce non è completamente noto. Attualmente stiamo conducendo esperimenti per verificare la nostra ipotesi secondo cui l'ANF influisce o sull'ingresso del calcio nelle cellule o sulla sua ridistribuzione all'interno delle cellule. L'ANF potrebbe esercitare questo effetto indirettamente, attivando la sintesi del guanosinmonofosfato ciclico (GMP ciclico). È questo un nucleotide che funge da secondo messaggero, cioè che trasmette all'interno della cellula il messaggio che è stato portato da un ormone fino alla superficie cellulare. Prova del rapporto tra ANF e GMP ciclico è il fatto che, nei ratti, l'iniezione del peptide porta a un significativo aumento dei livelli di GMP ciclico nel plasma e nell'urina. L'ANF potrebbe esercitare questo effetto attivando la guanilicociclasi, un enzima che è situato sulla membrana cellulare ed è interessato nella formazione del GMP ciclico. Inoltre, l'ANF inibisce la stimolazione dell'adenilicociclasi, l'enzima che è interessato nell'attività di quello che, in alcune cellule, è il secondo messaggero.

Esperimenti con ANF sintetico, marcato con isotopi radioattivi, hanno dimostrato che il peptide agisce sul cervello di ratti e cavie a vari livelli, legandosi ad aree responsabili della regolazione della pressione sanguigna e del controllo dei livelli di sodio, potassio e acqua. Essi hanno anche rivelato che l'ANF inibisce la produzione della vasopressina, un ormone che è sintetizzato nell'ipotalamo, situato nella parte basale del cervello, ed è poi trasportato nel lobo posteriore dell'ipofisi, dove viene accumulato. Quando la vasopressina è liberata dall'ipofisi, determina vasocostrizione e aumento della pressione sanguigna, e influenza il riassorbimento dell'acqua da parte dei tubuli renali. Infine, si è visto che l'ANF si lega a diversi siti nel corpo ciliare dell'occhio, probabilmente contribuendo così al controllo della pressione oculare.

Mentre prosegue lo studio degli effetti fisiologici dell'ANF, molti stanno indagando sulla possibilità di utilizzare questo ormone come farmaco per il controllo dell'ipertensione nell'insufficienza cardiaca di tipo congestizio. Noi abbiamo osservato una diminuzione significativa a breve termine (meno di un'ora) della pressione sanguigna in ratti ipertesi, trattati con ANF sintetico somministrato in un'unica iniezione di un microgrammo. La diminuzione più significativa è stata notata in animali in cui l'ipertensione dipendeva dall'attività della renina. Quando l'ANF è stato infuso per parecchi giorni, in dosi di un microgrammo all'ora, la pressione sanguigna è caduta in modo significativo, ritornando ai valori normali a partire dal secondo giorno. Un trattameno analogo, ma utilizzando dosi più piccole, somministrate per 12 giorni, ha portato alla caduta della pressione sanguigna negli ultimi 10 giorni e a un abbassamento dei livelli di aldosterone nel plasma e nell'urina.

Successivamente abbiamo misurato i livelli di ANF in un ceppo di ratti predisposti geneticamente all'ipertensione. Abbiamo trovato livelli alti di ANF in circolo e livelli inferiori alla norma nell'atrio sinistro (ma non in quello destro) del cuore. L'alto livello di ANF in circolo sembrerebbe un sintomo del tentativo dell'organismo di abbassare la pressione sanguigna, mentre i livelli più bassi nell'atrio sinistro sarebbero dovuti a deplezione dell'ANF che accompagna l'ipertensione.

L'ANF sembra avere anche un ruolo importante nell'insufficienza cardiaca congestizia. È questa una situazione in cui il cuore non riesce più a pompare il sangue correttamente; ne consegue che il paziente presenta fiato corto e pronunciato edema a livello delle gambe. Non si sa come l'ANF contribuisca alla malattia, ma studi su un ceppo di criceti affetti da insufficienza cardiaca congestizia spontanea hanno messo in relazione il cambiamento nei livelli di ANF con il progredire della malattia.

In questi criceti la pressione arteriosa si mantiene sempre su valori al di sotto della norma, mentre la pressione venosa aumenta con il progredire della malattia. A ogni stadio di questa, la quantità di ANF negli atri è inferiore ai valori riscontrati negli animali di controllo. Il risultato che colpisce di più è l'aumento significativo di ANF in circolo, rilevabile non appena la pressione venosa intracardiaca inizia ad aumentare. Il livello di ANF in circolo raggiunge un massimo quando la malattia è moderatamente avanzata, per poi diminuire negli stadi finali. Studi condotti *post mortem* sugli atri affetti hanno rivelato una «iperplasia da esaurimento», consistente in un aumento del reticolo endoplasmatico rugoso, che è la sede della sintesi del peptide, in un aumento delle dimensioni dell'apparato di Golgi, dove i peptidi vengono trasformati, e in una diminuzione del numero di granuli secretori, ciascuno dei quali contiene meno ANF del normale.

Questi risultati indicano che un modesto aumento della pressione atriale è sufficiente per stimolare una ipersecrezione dell'ANF. Un minor livello di ANF negli atri indica deplezione. Noi riteniamo anche che la ritenzione di sodio e la attivazione del sistema renina-angiotensina-aldosterone, due sintomi di insufficienza cardiaca congestizia, siano ritardate a stadi successivi della malattia grazie alla liberazione di grandi quantità di ANF. Il fatto che questi effetti compaiano alla fine può essere il risultato di una «regolazione da calo» (*down regulation*): in presenza di quantità molto forti di ANF, le cellule bersaglio hanno la possibilità di diminuire il numero di siti per l'ANF, presenti sulla loro membrana cellulare, con il risultato di rallentare le reazioni cellulari, che sono normalmente stimolate dal legame con l'ANF.

Ulteriori ricerche sull'ANF dovrebbero portare alla scoperta di nuovi sistemi di cura per l'ipertensione e per altre malattie che interessano la pressione sanguigna, per le malattie che alterano il volume ematico e per quelle renali che interferiscono sull'escrezione di sali e acqua. Tuttavia, nonostante i progressi nello studio delle malattie associate all'ANF, molto deve ancora essere fatto prima che forme sintetiche di questo peptide possano essere utilizzate in terapia. Resta tutta da chiarire la fisiologia degli effetti dell'ANF sui tubuli renali, come pure la connessione tra rilassamento dei vasi sanguigni, ridistribuzione del calcio all'interno delle cellule muscolari lisce ed effetti sulla guanilicociclasi e sulla adenilicociclasi. Occorre anche indagare sui fattori che stimolano la liberazione dell'ANF da parte dei cardiociti e descrivere il meccanismo con cui il peptide agisce sulle varie aree cerebrali. È positivo il fatto che sia stato confermato a livello clinico che tutti gli effetti dell'ANF riscontrati negli animali si ottengono anche nella specie umana. Ciò dovrebbe rendere più rapida la strada per utilizzare l'ANF nel trattamento delle malattie.

Lo sviluppo di una terapia basata sull'ANF dipenderà oltre che da una più approfondita comprensione dell'azione fisiologica del peptide, anche dallo sviluppo di tecniche per la produzione di suoi particolari analoghi, adatti per essere usati come farmaci nel trattamento delle singole malattie, agendo per ciascuna a livello di uno specifico sito di legame. Si dovrà inoltre riuscire a modificare questi farmaci in modo da proteggerli dall'attacco degli enzimi e degli acidi presenti nello stomaco, cosicché possano essere somministrati per via orale. La biotecnologia e la sintesi chimica renderanno disponibili le tecniche necessarie per questi scopi, ma occorreranno probabilmente molti anni prima che siano pronti analoghi dell'ANF da poter essere utilizzati per un'accurata sperimentazione sull'uomo.

Reazioni chimiche quantistiche in prossimità dello zero assoluto

Poiché si è visto che effetti quantomeccanici rendono possibili a bassissime temperature alcune reazioni chimiche normalmente «proibite», si pensa che le fredde e scure nubi di polvere galattica potrebbero contenere germi di vita

di Vitalii I. Goldanskii

In base alla nostra esperienza quotidiana sappiamo che le reazioni chimiche vengono rallentate dalle basse temperature. La carne, per esempio, può essere conservata mediante congelamento, una soluzione adottata anche presso i popoli antichi. Circa un secolo fa, il chimico fisico svedese Svante Arrhenius formulò una legge della chimica classica che stabilisce una relazione fra temperatura e velocità delle reazioni chimiche. Secondo l'equazione di Arrhenius, allo zero assoluto (zero kelvin, corrispondenti a -273 gradi centigradi o Celsius) la velocità di tutte le reazioni chimiche deve essere zero, il che significa in altre parole che le reazioni devono arrestarsi.

Le osservazioni sperimentali, tuttavia, dimostrano che l'equazione di Arrhenius, sebbene descriva generalmente con discreta precisione la velocità delle reazioni chimiche a temperature relativamente elevate, cessa di essere valida alle basse temperature. In tale dominio entra in gioco un effetto quantomeccanico noto come «effetto tunnel», o *tunneling*, che consente il verificarsi di reazioni chimiche non ammesse dal punto di vista classico. In particolare, interi atomi possono penetrare attraverso le barriere rappresentate dalle forze repulsive di altri atomi e costituire molecole complesse pur non essendo dotati, stando alle leggi della chimica classica, dell'energia necessaria per superare la repulsione di quegli atomi. Naturalmente, la velocità con la quale si formano le molecole complesse è estremamente bassa, ma il processo di tunneling può in ogni caso avere un ruolo significativo.

È in tale contesto che ho suggerito la possibilità di una preistoria fredda della vita: la formazione di molecole organiche sufficientemente complesse nel profondo freddo dello spazio esterno, dove le temperature raggiungono normalmente solo pochi kelvin. I raggi cosmici (protoni di alta energia e altre particelle) potrebbero innescare la sintesi di tali molecole in nubi scure di polvere interstellare. Successivamente le reazioni procederebbero, lentamente ma con continuità, sfruttando l'effetto tunnel. Circa un anno dopo la presentazione della mia ipotesi (1973), due astrofisici britannici, Fred Hoyle e N. C. Wickramasinghe, hanno affermato che molecole di formaldeide interstellare si sono effettivamente sviluppate formando polisaccaridi stabili come cellulosa e amido. Le loro conclusioni, sebbene aspramente discusse, hanno suscitato una certa eccitazione tra quei ricercatori che vorrebbero considerare le nubi galattiche come la sede ove si è svolta l'evoluzione prebiologica dei composti necessari alla vita.

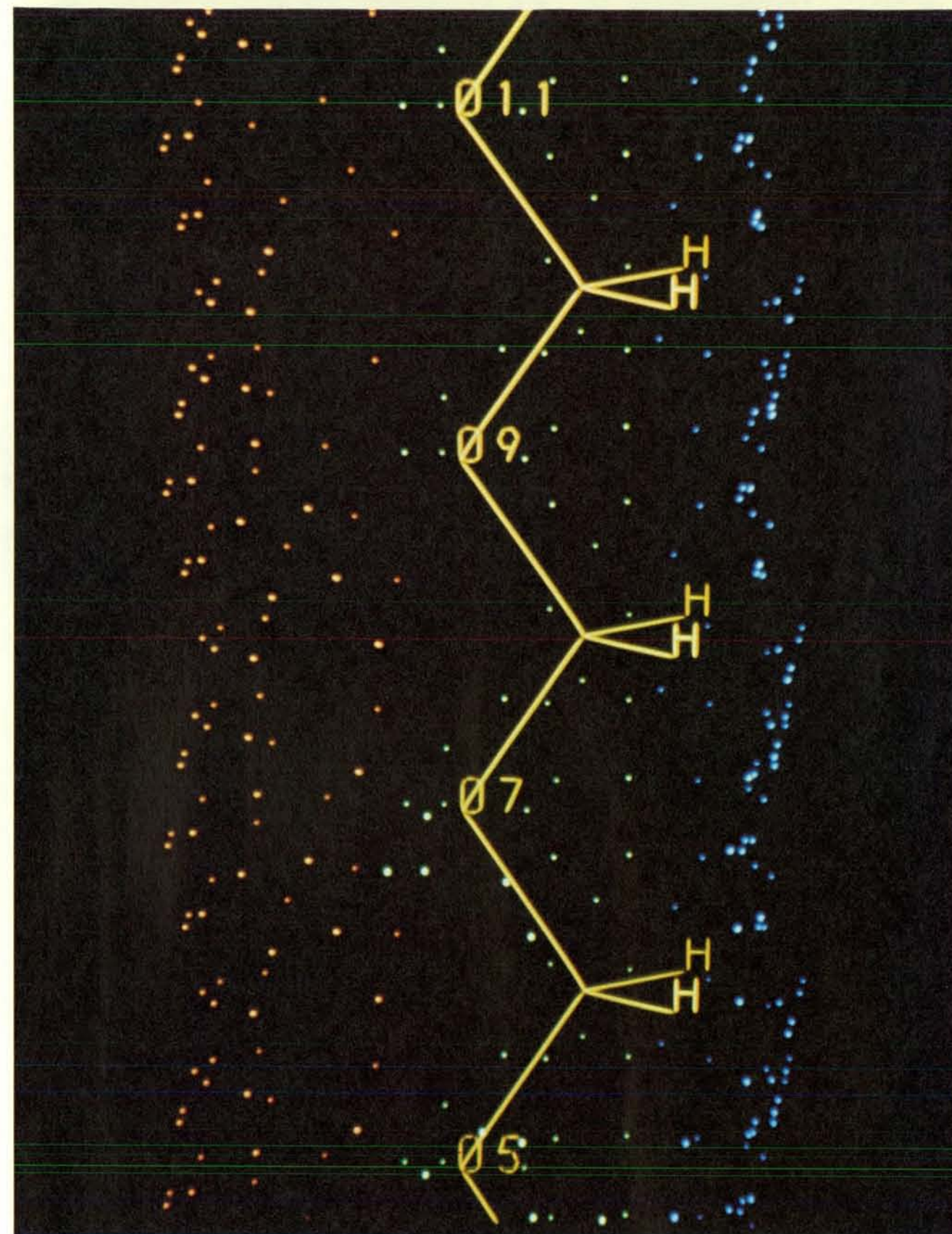
Non è mia intenzione sostenere o contraddire le asserzioni di Hoyle e Wickramasinghe. Desidero tuttavia far notare che, se la formaldeide interstellare si è sviluppata in molecole complesse, ciò può essere dovuto solo all'effetto tunnel. A proposito del ruolo svolto dalle reazioni chimiche a bassa temperatura nell'evoluzione di composti prebiologici, sembra appropriato richiamare le parole che il poeta russo Vladimir Mayakovski rivolgeva al poeta e romanziere Aleksandr Pushkin: «L'eternità è nostra, perché allora non spendere un paio d'ore?»

La velocità di una reazione chimica qualsiasi, in ambiente caldo o freddo, dipende da un fattore fondamentale noto come energia di attivazione della reazione, che può essere pensato nel modo seguente. Si immagini che ogni molecola sia sul fondo di una cosiddetta buca di potenziale energetico. La profondità e la larghezza della buca dipendono dalla disposizione spaziale degli atomi costituenti la molecola, che varia nel corso di una reazione chimica. Una reazione chimica corrisponde allora alla transizione di una molecola dal fondo di una buca di potenziale al fondo di un'altra. Secondo il punto di vista classico, una transizione di questo tipo può essere portata a termine solo superando una barriera di attivazione, ossia di potenziale, esistente tra le buche, l'altezza della quale viene detta energia di attivazione della reazione. Quanto maggiore è l'energia di attivazione, tanto minore è la velocità di reazione.

Di norma, le energie di attivazione sono 10 volte o più superiori all'energia termica delle molecole a temperatura ambiente, ovvero intorno ai 300 kelvin. Alla temperatura dell'elio liquido, che è di 4,2 kelvin, le energie di attivazione richieste sono oltre 1000 volte superiori all'energia termica delle molecole. Secondo la chimica classica, quindi, tanto alla temperatura ambiente, quanto a quella dell'elio liquido, molte reazioni non potranno aver luogo a meno che non venga fornita alle molecole una quantità supplementare di energia. In un certo senso la situazione è simile al meccanismo di una lotteria: prima che si possa vincere qualche cosa, occorre mettere insieme denaro sufficiente per comprare un biglietto della lotteria.

Si potrebbe tuttavia vincere, per così dire, senza neppure preoccuparsi di comprare un biglietto, se le molecole reagenti potessero aprirsi un varco e passare dal fondo di una buca di potenziale a una buca più profonda senza dovere scavalcare la barriera (energia di attivazione) che separa le due buche.

L'effetto tunnel è il diretto risultato di un principio fondamentale della teoria dei quanti: il dualismo onda-particella di ogni forma di materia e di radiazione. Nel macrocosmo visibile il dualismo, seppur non eliminato, viene trascurato e solitamente non è evidente. È nel microcosmo, dominio delle reazioni chimiche, che entrano in gioco gli effetti quantomeccanici. In particolare, il comportamento di «particelle», come atomi e molecole, è spesso determinato dalle loro caratteristiche ondulatorie.

Il tunneling quantistico di una particella

Il polimero della formaldeide, una lunga catena di singole molecole di formaldeide, può essere sintetizzato alla temperatura dell'elio liquido (4,2 kelvin, cioè -269 gradi centigradi o Celsius). Secondo le leggi classiche, a temperature così basse la maggior parte delle reazioni chimiche dovrebbe fermarsi; la sintesi dei polimeri della formaldeide è dovuta a un effetto quantomeccanico, noto come «effetto tunnel» o *tunneling*. Nel freddo profondo dello spazio esterno la formaldeide può anche trasformarsi in composti organici complessi come amido o cellulosa. Nell'immagine al calcolatore, elaborata da Paul Weiner e collaboratori della Rutgers University, i pallini rossi nei pressi degli atomi di ossigeno (O) indicano punti di potenziale negativo e quelli blu vicino agli atomi di idrogeno (H) e di carbonio (*vertici non contrassegnati*) punti di potenziale positivo.

attraverso una barriera di attivazione ha maggiore probabilità di accadere laddove lo spessore della barriera è inferiore alla lunghezza d'onda caratteristica della particella. Il fenomeno è analogo a ciò che accade quando un'onda luminosa colpisce la superficie di un materiale opaco. Parte dell'onda viene riflessa e parte viene assorbita dal mezzo. Se il mezzo opaco è abbastanza sottile, la parte d'onda che riesce ad attraversarlo apparirà sul lato opposto.

Uno dei primi successi della teoria dell'effetto tunnel si ebbe nel 1928, allorché George Gamow la applicò per spiegare quantitativamente i tassi di decadimento sperimentali di certi elementi radioattivi. Non più di un anno dopo David G. Bourgin dell'Università dell'Illinois ipotizzò che la barriera di attivazione delle reazioni chimiche potesse essere attraversata mediante un meccanismo di tunneling. Mentre lavorava all'Università di Oxford, Ronald P. Bell fornì validi contributi alla comprensione dell'effetto tunnel nelle reazioni chimiche. In particolare, egli dedusse alcune equazioni sulla probabilità del tunneling di particelle attraverso barriere di attivazione di conformazione variabile. Le sue previsioni sono state successivamente confermate per reazioni a temperature relativamente elevate, caratteristiche della fase liquida dei materiali.

In base all'equazione di Arrhenius, una legge della chimica classica, la sintesi dei polimeri richiederebbe alle basse temperature (*curva in nero*) una quantità di tempo sempre maggiore. Allo zero assoluto (zero kelvin) il tempo tenderebbe all'infinito e la reazione dovrebbe arrestarsi. Di fatto, però, la quantità di tempo (*curva in colore*) è finita; in altre parole, la reazione ha luogo, anche se procede lentamente. Come viene mostrato nell'illustrazione, il tempo necessario per aggiungere una nuova molecola alla catena in formazione si stabilizza a circa 10 millisecondi per temperature al di sotto dei 10 kelvin circa. Le misurazioni sono state eseguite nel laboratorio dell'autore.

In tempi più recenti l'attenzione dei ricercatori si è rivolta prevalentemente alle manifestazioni dell'effetto tunnel nelle reazioni chimiche a bassa temperatura. In queste condizioni, il numero delle transizioni quantistiche per effetto tunnel attraverso una barriera di attivazione supera il numero delle transizioni classiche di Arrhenius al di sopra della barriera. In un articolo del 1959 ho introdotto il concetto di «temperatura di tunneling»: la temperatura al di sotto della quale le transizioni «sotto» la barriera, cioè per effetto tunnel, sono molto più numerose delle transizioni «sopra» la barriera, ossia delle transizioni di Arrhenius, e la meccanica classica cede il passo alla sua controparte quantistica.

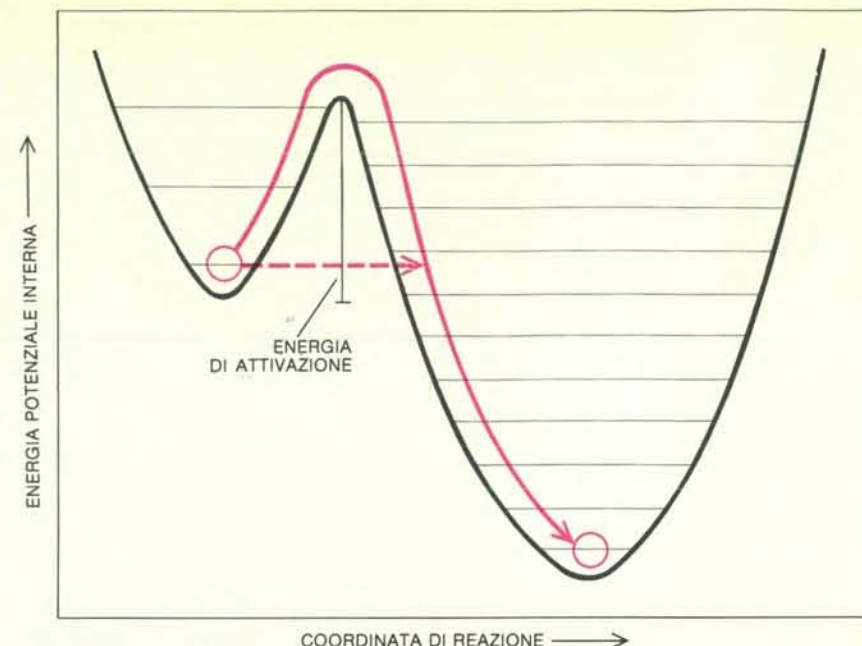
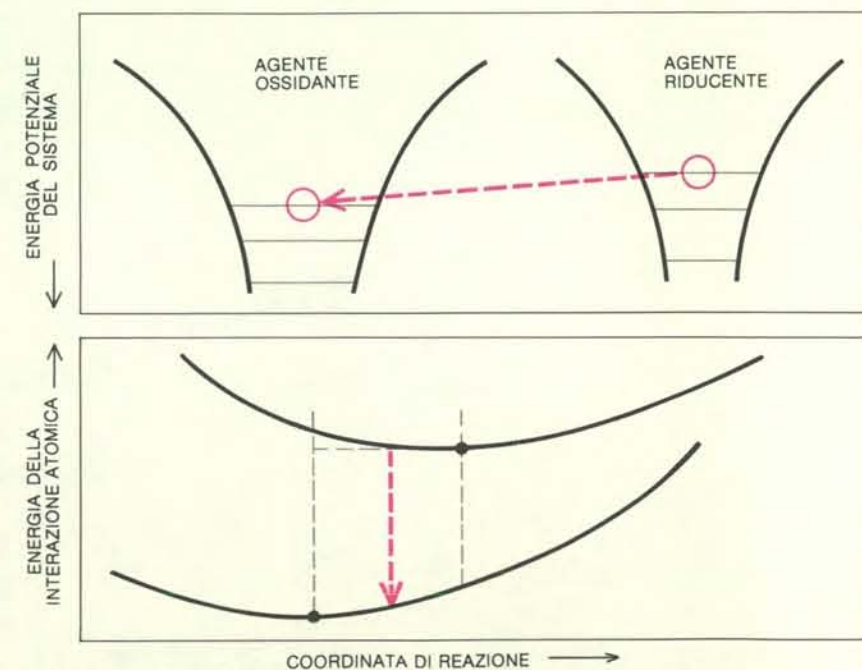
Si supponga, per esempio, che l'altezza di una barriera di attivazione sia pari a 0,4 elettronvolt (che è l'energia di attivazione caratteristica di molte reazioni chimiche) e che il suo spessore sia di 0,2 nanometri, vale a dire all'incirca il diametro di un atomo. (Un nanometro equivale a 10^{-9} metri.) La temperatura di tunneling calcolata per un atomo di idrogeno è quindi prossima a 160 kelvin e per il deuterio, l'isotopo pesante dell'idrogeno, è di circa 120 kelvin. Per atomi o molecole più pesanti la temperatura di tunneling è ancora più bassa, per il fatto che essa diminuisce proporzionalmente alla radice quadrata della massa della particella. Questo spiega perché, se si vogliono eseguire, nelle reazioni chimiche, osservazioni valide e significative dell'effetto tunnel, si dovrebbe operare nella regione criogenica (cioè in prossimità dello zero assoluto).

Alle basse temperature si manifesta anche un altro effetto. In prossimità dello zero assoluto tutte le molecole tendono al livello energetico di punto zero, cioè scendono alla base della barriera di attivazione, dove la distanza tra le buche di potenziale è massima. Dato, tuttavia, che la distanza rimane finita, la probabilità del tunneling raggiunge, al diminuire della temperatura, un livello finito che è indipendente dalla temperatura stessa. Quando ciò accade, la velocità di reazione tocca il proprio «limite quantistico a bassa temperatura». Nello stesso articolo del 1959, ho dedotto e analizzato l'esistenza di tale limite. Il tunneling attraverso la barriera energetica di attivazione non conduce di per sé necessariamente alla comparsa di un limite a bassa temperatura per la velocità di una certa reazione. Tuttavia l'esistenza di un limite indica che alle basse temperature gli effetti quantomeccanici prevalgono su quelli classici e che il tunneling esercita una funzione decisiva.

Come agisce l'effetto tunnel in quella importante classe di processi chimici noti come reazioni di ossidoriduzione? Questo tipo di reazioni, che ha interessato la chimica fin dalle sue origini, può essere schematicamente descritto come il processo di trasferimento di un elettrone da una specie chimica a un'altra. Un atomo o una molecola che cede elettroni è un agente riducente, o donatore di elettroni, mentre un atomo o una molecola che riceve elettroni è un agente ossidante, o accettore di elettroni. (In senso lato, un acido è un agente ossidante e una base è un agente riducente.)

Poiché la temperatura di tunneling decresce proporzionalmente alla radice quadrata della massa della particella che sfrutta l'effetto tunnel e la massa di un elettrone è circa 2000 volte inferiore a quella di un protone, la temperatura di tunneling di un elettrone dovrebbe, prevedibilmente, essere prossima a 7000 kelvin. Si potrebbe pertanto concludere che tutte le reazioni di ossidoriduzione procedono solo per effetto tunnel. Ma è facile spiegare perché le cose non stanno così. Diversamente dal libero moto degli elettroni nei metalli, nei superconduttori e negli altri conduttori, il moto degli elettroni nelle reazioni di ossidoriduzione è blandamente vincolato al moto dei nuclei atomici (molto più pesanti).

In questo processo lo spostamento delle particelle può essere paragonato al moto di un treno costituito da una leggera locomotiva agganciata a un vagone pesante con un giunto elastico e facilmente estensibile. L'elettrone è la locomotiva e il nucleo il vagone: anche se l'elettrone è in grado di percorrere una certa distanza prima che il nucleo cominci a muoversi, le due particelle sono legate. La temperatura di tunneling dell'elettrone dipende perciò sia dalla massa dell'elettrone, sia da quella del nucleo ed è quindi di norma sotto i 200 kelvin.

La prima inconfutabile dimostrazione della funzione svolta dall'effetto tunnel nelle reazioni di ossidoriduzione alle basse temperature è stata fornita nel 1966 da Britton Chance e Don C. DeVault della Università della Pennsylvania. Essi hanno ossidato (per sottrazione di elettroni) il citocromo *c*, riducendo (per addizione di elettroni) la clorofilla. Il citocromo *c* è una proteina trasportatrice di elettroni che, con altri citocromi, è disposta con un certo ordine lungo la catena respiratoria e va incontro alternativamente a ossidazione e riduzione, cedendo alla fine elettroni all'ossigeno molecolare. Chance e DeVault hanno rilevato che la velocità di reazione si stabilizzava a circa 120 kelvin e rimaneva costante fino alla temperatura più

Il tunneling tra buche di potenziale consente reazioni chimiche altrimenti impossibili. Le buche di potenziale corrispondono alle varie disposizioni spaziali degli atomi di una molecola. L'energia interna di ogni disposizione (*asse verticale*) varia lungo la coordinata di reazione, ossia con la distanza tra gli atomi della molecola (*asse orizzontale*). Si può immaginare che la molecola giaccia sul fondo di una buca. Poiché l'energia interna di una molecola cambia nel corso di una reazione chimica, si può rappresentare la reazione come la transizione di una molecola dal fondo di una buca di potenziale al fondo di un'altra. Tale transizione è ostacolata da una barriera, la cosiddetta energia di attivazione. Per la teoria classica, la molecola può spostarsi da una buca all'altra solo superando la barriera (*curva in colore*). Per la meccanica quantistica, invece, la molecola può attraversare la barriera (*freccia in colore tratteggiata*) anche a temperature molto basse. Le linee in grigio indicano i livelli energetici della molecola ammessi dalla teoria quantomeccanica.

Le reazioni redox (o di ossidoriduzione) consistono nel trasferimento di un elettrone da un agente riducente (un atomo o una molecola che cede elettroni) a un agente ossidante (un atomo o una molecola che riceve elettroni). L'illustrazione in alto mostra il tunneling di un elettrone di valenza, ossia esterno, da un agente riducente a un agente ossidante. Le buche di potenziale sono dovute all'attrazione esercitata dai nuclei delle due specie chimiche, con carica positiva, sull'elettrone dotato di carica negativa. L'illustrazione in basso rappresenta l'interazione atomica nell'agente riducente prima (*curva in alto*) che esso perda un elettrone e dopo (*curva in basso*). Il trasferimento dell'elettrone è indicato dalla freccia tratteggiata. Essendo esotermica, la reazione libera calore.

bassa da loro esaminata, 4,2 kelvin, che è la temperatura dell'elio liquido. Hanno poi calcolato che l'energia di attivazione era di 0,14 elettronvolt e che la distanza percorsa dall'elettrone per effetto tunnel era di circa tre nanometri, la dimensione lineare di un globulo proteico.

Presso l'Istituto di fisica chimica di Mosca, assieme ai miei colleghi Kiril I. Zamarayev, Alpha I. Mikhailov e Ravil F. Khairutdinov, ho preso in esame una vasta gamma di reazioni di ossidoriduzione che comportano il tunneling di elettroni tra ioni separati da distanze di parecchi nanometri. Abbiamo studiato la ricombinazione di elettroni e di anioni di ossigeno, O^-, dalla quale risulta la formazione di ioni O^{2-}. Una soluzione congelata di idrossido di sodio (NaOH) ionizzato forniva gli elettroni. Abbiamo innescato le reazioni irradiando la soluzione e abbiamo tenuto sotto controllo la formazione di O^{2-} applicando la risonanza di spin elettronico, o risonanza paramagnetica elettronica.

Per mezzo di questa tecnica, che è analoga alla risonanza magnetica nucleare, si controlla la risposta degli elettroni posti in un campo magnetico a frequenze di radiazione discrete. Abbiamo osservato che la velocità di formazione di ioni O^{2-} rimaneva costante nell'intervallo di temperature compreso tra 120 e 4,2 kelvin. L'invarianza della velocità è servita a dimostrare che la reazione procedeva per effetto tunnel. Abbiamo stimato che la distanza di tunneling fosse compresa tra tre e quattro nanometri. Abbiamo esaminato anche altri esempi più peculiari di tunneling in reazioni di ossidoriduzione.

Finora ho trattato solamente il caso del trasferimento di elettroni per mezzo dell'effetto tunnel. Tuttavia, una reazione chimica nel senso più completo del termine è un processo che, oltre a disporre gli atomi in maniera diversa, cambia anche la natura, la lunghezza e gli angoli dei legami di valenza che li uniscono.

L'esistenza di un limite quantistico di bassa temperatura per la velocità delle reazioni chimiche considerate sotto questa molteplicità di aspetti è stata da me scoperta assieme ai colleghi Igor M. Barkalov, Anatolii M. Kaplan, Dmitrii P. Kiryukhin, presso l'Istituto di fisica chimica di Mosca. Abbiamo studiato la formazione di lunghe catene di molecole da singole molecole di formaldeide (CH_2O). La formazione di queste catene è stata innescata sia con fasci di elettroni di alta energia provenienti da un acceleratore di particelle, sia con raggi gamma (radiazione elettromagnetica altamente energetica) emessi da atomi di cobalto 60 radioattivo. Una volta innescata, la crescita delle catene procede spontaneamente e non richiede un apporto supplementare di calore poiché le reazioni stesse sono esotermiche, ossia liberano calore.

Abbiamo operato entro un ampio intervallo di temperature criogeniche, da 140 a 4,2 kelvin. Misurando con un calorimetro la quantità di calore prodotta dalla reazione abbiamo determinato il calore di reazione. Abbiamo trovato che potevamo produrre lunghissime catene di molecole a partire dai monomeri, cioè dalle singole molecole, di formaldeide. A 140 kelvin le catene erano costituite da circa 10 milioni di unità monomeriche; a 77 kelvin, esse erano lunghe 100 000 unità e in prossimità di 4,2 kelvin la loro struttura comprendeva da 1000 a 2000 unità. A temperature superiori alla soglia approssimativa di 140 kelvin, il tempo medio necessario per l'aggiunta di un nuovo legame a una catena in formazione aumentava, al decrescere della temperatura, in accordo con la classica equazione di Arrhenius. A temperature sensibilmente inferiori a 140 kelvin, invece, l'incremento del valore di questo tempo medio si riduceva progressivamente fino a toccare, a 12 kelvin, il proprio limite quantistico a bassa temperatura: un centesimo di secondo. A 4,2 kelvin la velocità di reazione era di 110 ordini di grandezza superiore a quanto avrebbe fatto prevedere l'equazione di Arrhenius.

I nostri risultati ben si adattano alla teoria secondo la quale la reazione procede sfruttando il tunneling delle singole molecole di formaldeide. Una molecola si trasferisce da una buca di potenziale relativamente poco profonda, prodotta dalla sua interazione con le molecole circostanti, a

La velocità di una reazione di ossidoriduzione è stata misurata nel caso specifico della ricombinazione di elettroni con anioni di ossigeno, O^-, dalla quale vengono prodotti anioni O^{2-} (specie dotata di doppia carica negativa). Una soluzione congelata di idrossido di sodio (NaOH) ionizzato è servita da donatore di elettroni. Poiché il tasso di decremento della concentrazione di elettroni (*curva in nero*) era pari al tasso di decremento della concentrazione di anioni O^- (*curva in colore*), si è potuto concludere che nel corso della reazione si formavano anioni O^{2-}. (Le concentrazioni dei reagenti sono normalizzate in base ai loro valori iniziali; le scale sull'asse verticale sono leggermente sfalsate per evitare sovrapposizioni.) Simboli differenti corrispondono a misurazioni che sono state eseguite a temperature differenti, comprese nell'intervallo tra 120 kelvin e 4,2 kelvin. Il fatto che la velocità di ricombinazione sia rimasta costante lungo tutto questo intervallo di temperature indica che la reazione procedeva sfruttando l'effetto tunnel; secondo la teoria chimica classica, la velocità della reazione dovrebbe diminuire con l'abbassarsi della temperatura.

una posizione contigua associata a una buca di potenziale più profonda. Questa molecola va a costituire il legame terminale della catena di monomeri di formaldeide, ove essa compare in una posizione energeticamente più stabile, accompagnata dalla liberazione e dalla successiva dissipazione di 0,4 elettronvolt di energia.

La crescita di catene di unità di formaldeide per effetto tunnel può essere paragonata alla fissione spontanea dei nuclei atomici. Secondo quanto rivela l'analisi teorica dell'effetto tunnel, la probabilità che la barriera di attivazione venga penetrata è proporzionale al prodotto tra lo spessore della barriera, la radice quadrata dell'altezza della barriera e la radice quadrata della massa della particella che la attraversa. La massa di una specie chimicamente reattiva è paragonabile alla massa di un frammento di fissione; l'altezza della barriera relativa a una comune reazione chimica è circa 10 milioni di volte inferiore a quella della barriera relativa alla fissione nucleare, mentre lo spessore della barriera è nel primo caso quasi 3000 volte maggiore. Gli effetti dell'altezza e dello spessore, pertanto, si compensano reciprocamente (la radice quadrata di 10 milioni è approssimativamente 3000) e la probabilità di tunneling della formaldeide è, abbastanza imprevedibilmente, paragonabile alla probabilità della fissione nucleare.

Il tunneling molecolare di un monomero di formaldeide (*in alto a destra*) può portare alla formazione di una catena polimerica (*in alto a sinistra*). La formaldeide è costituita da un atomo di ossigeno (O) e da due atomi di idrogeno (H), non indicati, tutti e tre legati a un atomo di carbonio. Il complesso al centro della sequenza in alto riproduce lo stadio intermedio del processo: la rottura del doppio legame tra atomi di ossigeno e atomi di carbonio che consente la formazione dei legami singoli —C—O—C—O— del polimero. In basso sono mostrate le buche di potenziale relative a ogni stadio della polimerizzazione. Ripetendo il processo, si aumenta la lunghezza del polimero.

Il quadro reale è più complesso di quello che ho delineato in questo articolo: la posizione di una singola molecola di formaldeide prossima all'estremità libera di una catena in formazione non è rigidamente fissa e la molecola può oscillare tra posizioni immediatamente adiacenti. L'aggiunta di una nuova molecola alla catena avviene molto probabilmente nel momento in cui la penultima molecola della catena si trova il più vicino possibile all'estremità della catena. In quell'istante l'altezza e lo spessore della barriera di attivazione hanno il valore minimo. Quando, invece, la penultima molecola assume una posizione lontano dall'estremità della catena, la probabilità di acquisire un nuovo legame è estremamente bassa. Sia l'altezza sia lo spessore della barriera di attivazione sono in questo caso molto grandi ed è molto minore la frequenza sia delle transizioni che superano la barriera sia delle transizioni che la attraversano.

Gli effetti di queste oscillazioni intermolecolari relativamente lente sono stati chiaramente osservati negli esperimenti eseguiti dal nostro gruppo all'Istituto di fisica chimica di Mosca. Abbiamo preso in esame la reazione innescata da radiazioni dell'acido bromidrico con l'etilene. Si è osservato che la sostituzione dell'idrogeno con il deuterio riduce la velocità di reazione a un quarto del valore normale a temperature prossime alla temperatura di tunneling. Se invece la reazione si fosse svolta tra molecole stazionarie, il fattore di decremento sarebbe stato prossimo a 100 000.

L'effetto tunnel mitigato da moti intermolecolari può intervenire in misura con-

I limiti della velocità di reazione alle basse temperature indicano che si è avuto un tunneling. Limiti siffatti sono stati osservati per molte reazioni chimiche: (*1*) la formazione di catene polimeriche di formaldeide (1973); (*2*) il legame dell'ossido di carbonio con l'emoglobina (1975); (*3*) l'isomerizzazione, cioè la ridisposizione strutturale, di coppie di radicali nella dimetilgliossima in seguito all'irradiazione con raggi gamma (1977); (*4, 5*) la rimozione di atomi di idrogeno da metanolo (*4*) ed etanolo (*5*) congelati mediante radicali metilici (1977); (*6*) il trasferimento di atomi di idrogeno durante l'isomerizzazione di certi radicali (1978); (*7*) la formazione di legami carbonio-carbonio in certi radicali (1979); (*8*) la bromurazione (aggiunta di acido bromidrico) dell'etilene (1978); (*9*) la clorurazione del cloruro di butile mediante cloro molecolare (1980). Le osservazioni di cui si riportano i dati sono state effettuate in Unione Sovietica, Stati Uniti, Giappone e Canada.

Particolari tipi di calorimetro, uno strumento che misura la quantità di calore prodotta o assorbita nel corso di una reazione, sono stati impiegati nel laboratorio dell'autore per determinare la velocità di polimerizzazione della formaldeide alle temperature criogeniche. Lo strumento a sinistra si è dimostrato efficace per temperature comprese tra 350 e 80 kelvin. Dall'alloggiamento del calorimetro (*a sinistra*) sporge un contenitore di vetro contenente monomeri di formaldeide. Durante gli esperimenti, l'intera unità era posta in un contenitore più grande e raffreddata con azoto liquido. Il calorimetro a destra è stato impiegato per temperature fino a 4,2 kelvin. La parte più esterna di questo strumento veniva riempita di azoto liquido, mentre la parte più interna conteneva elio liquido. Entrambi i calorimetri sono stati ideati da Igor M. Barkalov, Anatolii M. Kaplan e Dmitrii P. Kiryukhin dell'Istituto di fisica chimica di Mosca. L'intero procedimento sperimentale è stato seguito e controllato dall'autore.

siderevole anche nell'attività delle proteine e del DNA. Se ne possono avvertire le conseguenze non soltanto alle temperature criogeniche, ma anche a temperature superiori a 200 kelvin e persino a temperature fisiologiche. In particolare, l'effetto tunnel può contribuire in misura significativa alle transizioni fra i vari sottostati conformazionali delle molecole dei polimeri biologici.

I sottostati conformazionali corrispondono a configurazioni leggermente differenti di una stessa molecola le quali sono però dotate di energie più o meno equivalenti. Un atomo o un gruppo di atomi può, per esempio, essere ruotato intorno a un singolo legame. Le transizioni da un sottostato all'altro spiegano la mobilità delle molecole dei biopolimeri e possono anche essere implicate nella loro specifica azione. Hans Frauenfelder, con i colleghi dell'Università dell'Illinois a Urbana-Champaign e della Wayne State University, e David Phillips e collaboratori dell'Università di Oxford hanno scoperto e analizzato le proprietà dei sottostati conformazionali. Entrambi i gruppi hanno impiegato tecniche ausiliarie quali, per esempio, l'analisi per diffrazione di raggi X. Con Rudolf Mössbauer e Fritz Parak della Technische Universität di Monaco di Baviera, anche i miei collaboratori e io abbiamo analizzato i sottostati conformazionali.

Lo studio delle transizioni fra sottostati conformazionali ha condotto all'elaborazione di nuovi modelli dinamici delle proteine e di altri polimeri biologici. Circa 40 anni fa l'eminente fisico austriaco Erwin Schrödinger, nel suo libro *Scienza e umanesimo. Che cos'è la vita?* (Sansoni, Firenze), postulava che le proteine fossero cristalli aperiodici, ovvero irregolari. Circa una decina di anni fa alcuni ricercatori avanzarono l'ipotesi secondo la quale una proteina è costituita da un nucleo quasi solido circondato da un guscio di natura fluida. Secondo questo punto di vista, il guscio rappresenta la maggior parte del volume del globulo proteico. La terza e più recente teoria afferma che proteine e DNA possono essere considerati alla stregua di solidi vetrosi non omogenei e che le transizioni fra i loro sottostati alle temperature criogeniche e forse anche oltre i 200 kelvin sono effettivamente esempi di tunneling.

Due osservazioni finali appaiono necessarie su alcuni effetti del tunneling. Sebbene svolga una funzione importante nella formazione di legami chimici, il tunneling può anche contribuire alla distruzione di legami, e quindi di composti chimici. Esso, infatti, favorisce la penetrazione attraverso barriere di attivazione che tendono a opporsi ai processi distruttivi.

Il fatto che l'effetto tunnel consenta reazioni chimiche a temperature molto basse aggiunge un ulteriore motivo di dubbio circa l'eventualità di far rivivere organismi complessi che siano stati per lungo tempo ibernati. Quanto meno tali organismi dovrebbero essere stati protetti da eventi che innescano reazioni chimiche, come, per esempio, da radiazioni penetranti.

I dati sperimentali e le considerazioni teoriche concernenti le manifestazioni dell'effetto tunnel nelle reazioni chimiche, come pure le numerose applicazioni della chimica quantistica in prossimità dello zero assoluto, testimoniano a favore del fatto che la chimica non è decaduta in condizioni servili rispetto alla fisica e alla biologia. Essa offre, ora come sempre, una proficua collaborazione nello scambio reciproco di idee nuove e fondamentali.

Confronto tra DNA e filogenesi degli uccelli

Le diversità tra acidi desossiribonucleici mettono in luce le distanze evolutive che separano le specie alle quali essi appartengono e consentono di ricostruire e di datare le ramificazioni dell'albero evolutivo degli uccelli

di Charles G. Sibley e Jon E. Ahlquist

Tutti gli esseri viventi hanno antenati e, di conseguenza, tutti gli esseri viventi hanno una storia evolutiva. Poiché, probabilmente, sia le piante sia gli animali hanno avuto origine da un unico capostipite i membri di ciascuno dei due gruppi hanno una medesima storia o filogenesi: la ricostruzione della filogenesi è uno dei principali scopi della biologia evoluzionistica. Le specie attualmente viventi occupano i ramoscelli estremi di un grande albero filogenetico in cui rami e tronco non sono più visibili da lungo tempo; per ricostruire l'albero della vita è necessario scoprirne lo schema di ramificazione e, se possibile, datare gli eventi del passato.

Negli ultimi dieci anni abbiamo utilizzato una tecnica, l'ibridazione DNA-DNA, che consente di trarre informazioni di tipo filogenetico direttamente dal materiale genetico, e l'abbiamo applicata alla ricostruzione della filogenesi degli uccelli. Gli uccelli contano oggi circa 9000 specie viventi, che costituiscono la discendenza di linee evolutive che hanno iniziato a divergere l'una dall'altra qualcosa come 150 milioni di anni fa, alla fine del Giurassico o all'inizio del Cretaceo, immediatamente dopo l'origine degli uccelli da un antenato di tipo rettiliano. In base alla stima di Pierce Brodkorb, dell'Università della Florida, almeno 150 000 specie di uccelli sono esistite da allora a oggi e, di queste, le specie viventi costituiscono appena il 6 per cento, mentre tutte le altre sono ormai estinte.

Nella nostra ricerca abbiamo provato a misurare le differenze medie tra DNA di specie rappresentative dei principali gruppi di uccelli viventi, e abbiamo tentato di ricostruire, sulla base dei risultati, la sequenza delle ramificazioni nell'albero evolutivo degli uccelli. La filogenesi così ricostruita fornisce la base per una classificazione degli uccelli in cui le specie viventi sono assegnate alle varie categorie tassonomiche per le loro relazioni genealogiche; in qualche caso, i nostri risultati differiscono dalla nota classificazione tradizionale.

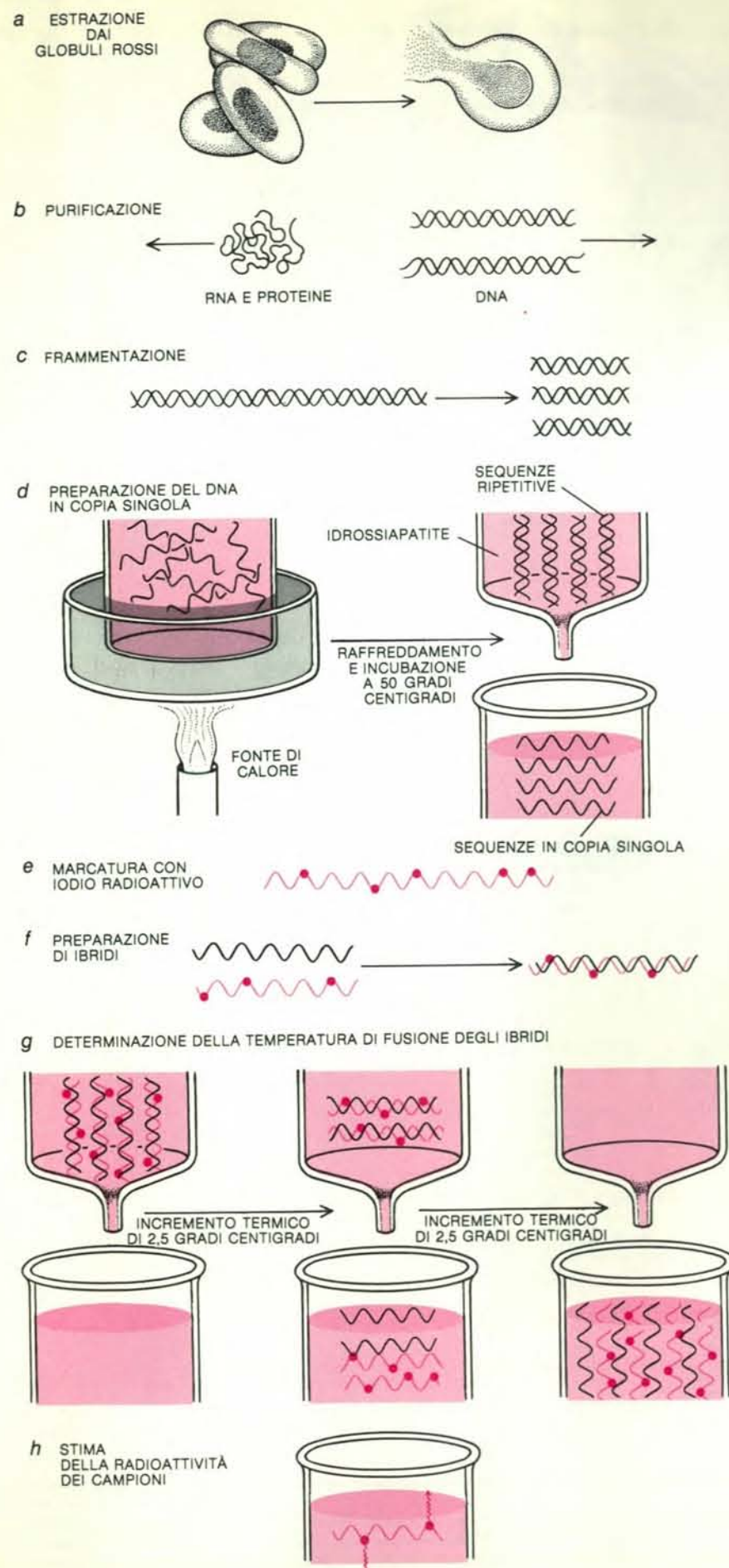
Gli elementi necessari per ricostruire una filogenesi sono lo schema di ramificazione e la datazione di ciascun evento che ha portato a quest'ultima: eventi del genere avvengono quando una barriera, solitamente di tipo geografico, divide una specie in due popolazioni, che possono così divergere geneticamente e diventare capostipiti di due diverse linee evolutive. Il processo di biforcazione può ripetersi più volte in ciascuna linea filetica, dando luogo a una varietà di specie morfologicamente ed ecologicamente differenti (radiazione).

Finora, l'unica fonte di informazioni circa i rapporti filogenetici tra specie era il confronto di caratteri morfologici e anatomici, confronto che ha dato una risposta a molti interrogativi e ha contribuito ad abbozzare a grandi linee la storia dei viventi. I caratteri anatomici, però, si sono modellati in risposta a precise esigenze funzionali; non è detto, quindi, che una somiglianza morfologica indichi necessariamente relazioni di parentela, perché forme simili possono insorgere, per convergenza evolutiva, anche in organismi non imparentati tra loro. Rondini e rondoni, per esempio, sono molto simili a una osservazione superficiale, il che è dovuto al fatto che entrambi i gruppi sono specializzati nella predazione di insetti alati: essi sono tanto simili che, nelle prime classificazioni degli uccelli, venivano considerati insieme. Studi successivi, riscontrando differenze anatomiche fondamentali, hanno dimostrato invece che i due gruppi non sono affatto strettamente affini, dato che i rondoni sono lontani parenti dei colibri, mentre le rondini sono tipici passeriformi. Molti casi di convergenza sono così sottili che diventa molto difficile risolverli sulla sola base dei confronti morfologici.

Per datare gli eventi di ramificazione dedotti dai confronti anatomici è stato necessario rifarsi ai reperti fossili. Le filogenesi parziali di alcuni gruppi sono state così ricostruite sulla base di dati paleontologici, ma per molti altri gruppi, tra cui gli uccelli, la documentazione fossile è frammentaria. Inoltre, se è vero che un fossile datato indica, con una certa approssimazione, l'epoca in cui un certo organismo è vissuto e morto, è altrettanto vero che esso non ci dice nulla sul momento in cui la sua linea filetica ha cominciato a divergere da quella degli altri organismi imparentati.

Indubbiamente, una ricostruzione filogenetica risulterebbe molto più corretta se potesse utilizzare un metodo che consentisse di misurare direttamente le distanze genealogiche tra le linee filetiche ancora esistenti, e di datare i tempi in cui si sono avute divergenze. Le relazioni genetiche tra le specie viventi riflettono la loro evoluzione e, poiché i mutamenti genetici sono perloppiù divergenti, la distanza genetica tra due linee filetiche qualsiasi è legata al tempo trascorso da quando esse hanno avuto l'ultimo antenato comune. Per ricostruire la filogenesi degli uccelli ne abbiamo quindi studiato il materiale genetico.

In tutti gli organismi, a eccezione di alcuni virus, il materiale genetico è il DNA, una molecola composta da due filamenti, ciascuno costituito da una sequenza di quattro tipi di unità chimiche, i nucleotidi. Ogni nucleotide a sua volta è composto da un pentoso (zucchero a cinque atomi di carbonio), da un gruppo fosfato e da una base azotata. I vari nucleotidi differiscono l'uno dall'altro solo per la base azotata: adenina (A), timina (T), citosina (C) e guanina (G). Le basi, lungo i due filamenti formano coppie complementari, tramite legami a idrogeno: A si unisce sempre a T e C a G. L'informazione genetica è codificata nella sequenza delle basi e sequenze specifiche di basi formano i geni, che dirigono la sintesi di molti tipi di proteine, i quali partecipano alla costruzione della maggior parte delle strutture animali e vegetali e ne controllano le funzioni.

I rami sui quali barbuti (capitoni) e tucani sono appollaiati simboleggiano le relazioni filogenetiche tra i gruppi ai quali queste specie appartengono. I barbuti del Nuovo Mondo e quelli africani e asiatici sono tradizionalmente collocati in un'unica famiglia, i tucani in un'altra. I confronti tra i DNA hanno però dimostrato che, contrariamente a quanto sembrerebbe dall'aspetto, i barbuti del Nuovo Mondo sono più strettamente imparentati con i tucani che con i barbuti del Vecchio Mondo. L'entità delle differenze osservate tra i DNA indica che la divergenza tra i due gruppi di barbuti è iniziata circa 55 milioni di anni fa, mentre tucani e barbuti del Nuovo Mondo si sono separati circa 30 milioni di anni fa.

All'interno del genoma di ciascuna cellula (il corredo completo dei geni di quella cellula), la maggior parte dei geni è presente in copia singola. Negli organismi superiori alcune sequenze (una percentuale che va dal 3 al 5 per cento) sono presenti, invece, in più di una copia e queste sequenze ripetitive possono costituire anche il 40 per cento del volume di DNA di una cellula.

Nel genoma di un uccello vi sono circa due miliardi di coppie di nucleotidi: la tecnica dell'ibridazione DNA-DNA consente di confrontare tra loro anche quantità così sterminate di unità genetiche e di misurare, in tal modo, le differenze genetiche tra le specie oggi esistenti. Queste misurazioni permettono di ricostruire la struttura dell'albero filogenetico e anche, riportando le differenze genetiche su una scala temporale opportunamente calibrata, di calcolare con una certa approssimazione i tempi di divergenza tra le linee filetiche attuali.

La tecnica dell'ibridazione si basa sulle proprietà delle molecole di DNA. Infatti, se una soluzione contenente DNA in filamento doppio è riscaldata fino all'ebollizione, i legami a idrogeno tra basi complementari «fondono», o si rompono, e il DNA si separa in filamenti singoli. I legami a idrogeno sono i legami chimici più deboli presenti nel DNA; il resto della molecola non è danneggiato dal calore. Quando il campione fuso si raffredda, i filamenti singoli si scontrano a caso e, se possiedono sequenze complementari di basi, si riassociano nella struttura a filamento doppio, dato che queste basi si «riconoscono» a vicenda e possono ristabilire tra loro legami a idrogeno. Se la riassociazione avviene a bassa temperatura, il doppio filamento riformatosi (duplex) può contenere numerose coppie sbagliate, ma a 60 gradi centigradi (o gradi Celsius), perché il duplex sia stabile, occorre che almeno l'80 per cento delle basi formi coppie giuste. In queste condizioni, i filamenti singoli si riassociano solo con i loro complementari e si restaura così il doppio filamento originario di DNA.

Se filamenti singoli di DNA, provenienti da due specie diverse, vengono posti insieme in incubazione a 60 gradi centigradi, si formeranno doppi filamenti ibridi solo tra sequenze omologhe di basi, cioè tra sequenze ereditate da un antenato comune alle due specie; infatti, solo sequenze omologhe contengono un numero di basi complementari sufficiente a rendere l'ibrido termicamente stabile a 60 gradi centigradi. Un duplex ibrido di DNA provenienti da specie diverse conterrà coppie sbagliate di basi, perché nelle due diverse linee filetiche, a partire dall'ultimo antenato comune, si saranno accumulate differenti serie di mutazioni. Così, una *A* potrà trovarsi di fronte a una *C*, o una *G* di fronte a una *T*, e tra queste coppie non si formeranno legami a idrogeno. Siccome la temperatura di fusione del duplex è proporzionale al numero di legami a idrogeno, la presenza di coppie sbagliate farà sì che il DNA ibrido fonda a una temperatura più bassa di quella necessaria alla fusione di un filamento doppio perfettamente appaiato.

L'ibridazione DNA-DNA richiede una fonte cellulare di questo acido nucleico. Nel caso degli uccelli, il DNA viene estratto rompendo i globuli rossi del sangue, che in questi animali sono nucleati (*a*). Esso viene poi separato dall'RNA e dalle proteine (*b*), e i suoi lunghi filamenti vengono tagliati in frammenti più corti (*c*). A questo punto, viene portato a ebollizione per breve tempo in modo che i doppi filamenti si separino in filamenti singoli. Quando il DNA si raffredda, le sequenze ripetitive, presenti in numerose copie, si riassociano più rapidamente di quelle presenti in copia singola. Il DNA riassociato viene allora fatto passare attraverso una colonna di idrossiapatite, che lega i filamenti doppi di DNA, ma non quelli singoli (*d*). In questo modo, molte delle sequenze ripetitive, che hanno due filamenti, si legano all'idrossiapatite, mentre i frammenti di DNA contenenti copie singole di ciascun gene passano oltre. Il DNA in copia singola viene quindi marcato con iodio radioattivo (*e*) e mescolato con DNA non marcato della stessa specie o di una specie diversa. Il miscuglio di DNA diversi viene incubato a 60 gradi centigradi per 120 ore; in esso si forma un DNA ibrido contenente un filamento marcato e uno no (*f*). Ogni DNA ibrido viene posto su una colonna di idrossiapatite, che viene riscaldata a bagnomaria tra 55 e 95 gradi centigradi con incrementi successivi di 2,5 gradi centigradi. A ogni incremento termico, i singoli filamenti derivanti dalla fusione degli ibridi vengono eluiti dalla colonna e raccolti in un flacone (*g*). La misura della radioattività presente nei diversi flaconi indica quanta parte dell'ibrido è andata incontro a fusione a ogni temperatura (*h*).

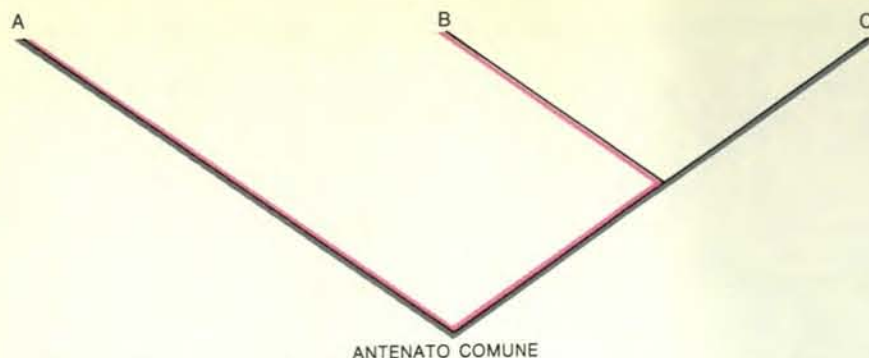
DNA MARCATO CON IODIO RADIOATTIVO
DNA NON MARCATO
MISCELAZIONE, FUSIONE E INCUBAZIONE
a DOPPI FILAMENTI NON RADIOATTIVI
b DNA IN FILAMENTO SINGOLO
c DOPPI FILAMENTI RADIOATTIVI
d DUPLEX IBRIDI
ZONA CONTENENTE APPAIAMENTI ERRATI
COMPLEMENTARITÀ ELEVATA
COMPLEMENTARITÀ BASSA

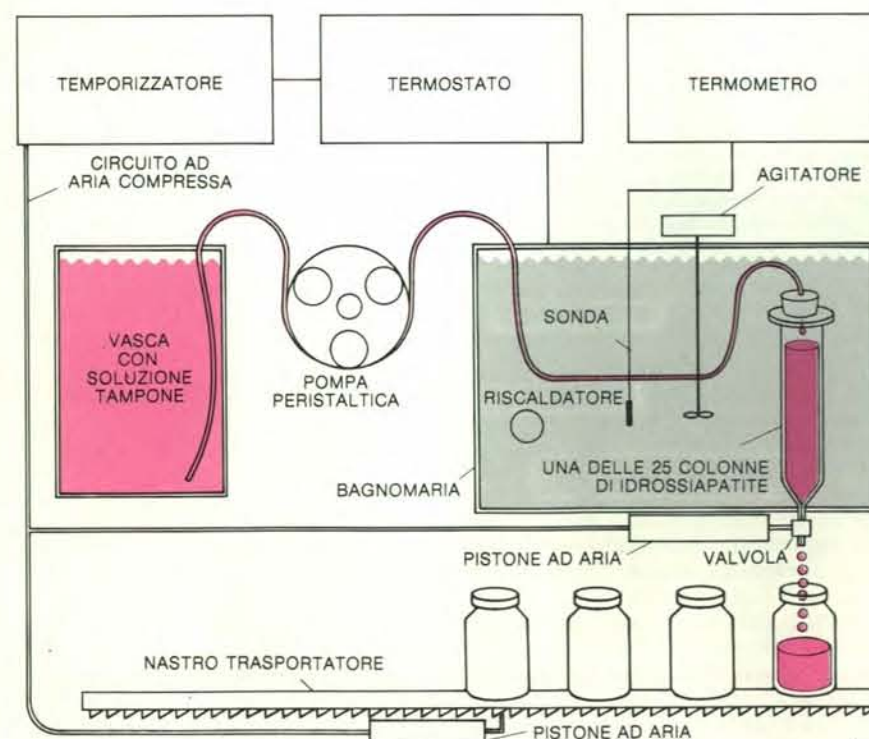
Per formare i DNA ibridi si mescolano una piccola quantità di DNA radioattivo e una quantità 1000 volte maggiore di DNA non marcato, della stessa specie o di una specie diversa (*in alto*). Il DNA viene scaldato per separare il doppio filamento in filamenti singoli, poi incubati per favorirne la riassociazione. Buona parte del DNA non marcato si riassocia con filamenti complementari non marcati (*a*), dando luogo a doppi filamenti che non sono radioattivi e perciò non modificano le successive misurazioni. Alcuni dei filamenti singoli non si riassociano (*b*) e solo l'1 per cento circa del DNA marcato si ibrida con altri filamenti marcati (*c*). Una parte del DNA forma duplex ibridi, composti da un filamento marcato e da uno non marcato (*d*). In ibridi di DNA di due specie diverse, la percentuale delle basi di un filamento che sono accoppiate con basi complementari del filamento opposto è in funzione della somiglianza genetica tra le due specie. Poiché molti più legami tengono uniti i due filamenti negli ibridi con basi ben accoppiate (*in basso a sinistra*), questi ibridi fondono a temperature più elevate rispetto agli ibridi con molte basi non ben accoppiate (*in basso a destra*).

Il DNA per l'ibridazione DNA-DNA viene estratto dai nuclei cellulari e separato dalle proteine e dagli altri componenti della cellula. I filamenti vengono quindi tagliati in frammenti mediamente di 500 nucleotidi. La maggior parte delle sequenze ripetitive viene poi rimossa dal DNA della specie che si vuol confrontare con altre e il restante DNA in «copia singola» viene marcato con iodio radioattivo. Una piccola quantità di DNA radioattivo, nota come «tracciante» (*tracer*) viene combinata con una quantità molto maggiore di DNA non marcato, l'«elemento guida» (*driver*) della medesima specie. Uno stesso tracciante può anche essere combinato con elementi guida di altre specie. Ciascun miscuglio produrrà un suo tipo di ibrido: un omoduplex nel caso in cui la specie dalla quale sono stati ottenuti il tracciante e l'elemento guida sia la stessa, o un eteroduplex, nel caso in cui i due filamenti provengano da specie diverse. L'omoduplex costituisce lo standard di riferimento con cui si confronta la stabilità termica degli eteroduplex.

I miscugli «tracciante-elemento guida» vengono fatti bollire per cinque minuti in modo da dissociare i filamenti doppi in filamenti singoli. Per favorire la riassociazione in duplex, i filamenti singoli vengono fatti incubare per 120 ore in una soluzione tampone di fosfato di sodio a 60 gradi centigradi. Ognuno dei differenti campioni di DNA ibrido a duplice filamento, così formati, viene posto su una colonna di idrossiapatite, una forma cristallina di fosfato di calcio che ha la proprietà di legare il DNA a filamento doppio, ma non quello a filamento singolo. Le colonne sono poi immerse in acqua, in un bagnomaria a 55 gradi centigradi, la cui temperatura viene portata, per incrementi successivi di 2,5 gradi, fino a 95 gradi centigradi. A ciascuna delle 17 temperature i frammenti di DNA a filamento singolo, originatisi per fusione dei duplex, vengono asportati dalla colonna e raccolti in un flacone. La radioattività presente in ciascun flacone viene misurata e indica così quanta parte dell'ibrido corrispondente si è fusa. I risultati permettono di tracciare una curva di fusione, cioè un grafico che mostri la quantità di ibrido che si è fusa a ciascuna temperatura. La differenza media, in gradi centigradi, tra la curva dell'omoduplex e quella di ciascun eteroduplex è una misura della differenza genetica media tra la specie da cui è stato estratto il tracciante e la specie di ogni elemento guida, confrontata.

La differenza tra i DNA di due specie può essere utilizzata come indicatore della distanza genealogica tra le due specie, solo se si può postulare che il DNA muti con una velocità media uguale in qualsiasi linea filetica. L'ipotesi che l'evoluzione delle proteine procede a velocità costante è stata formulata nel 1962 da Emile Zuckerkandl e Linus Pauling, allora al California Institute of Technology, ed è da allora che si discute di «orologi molecolari».

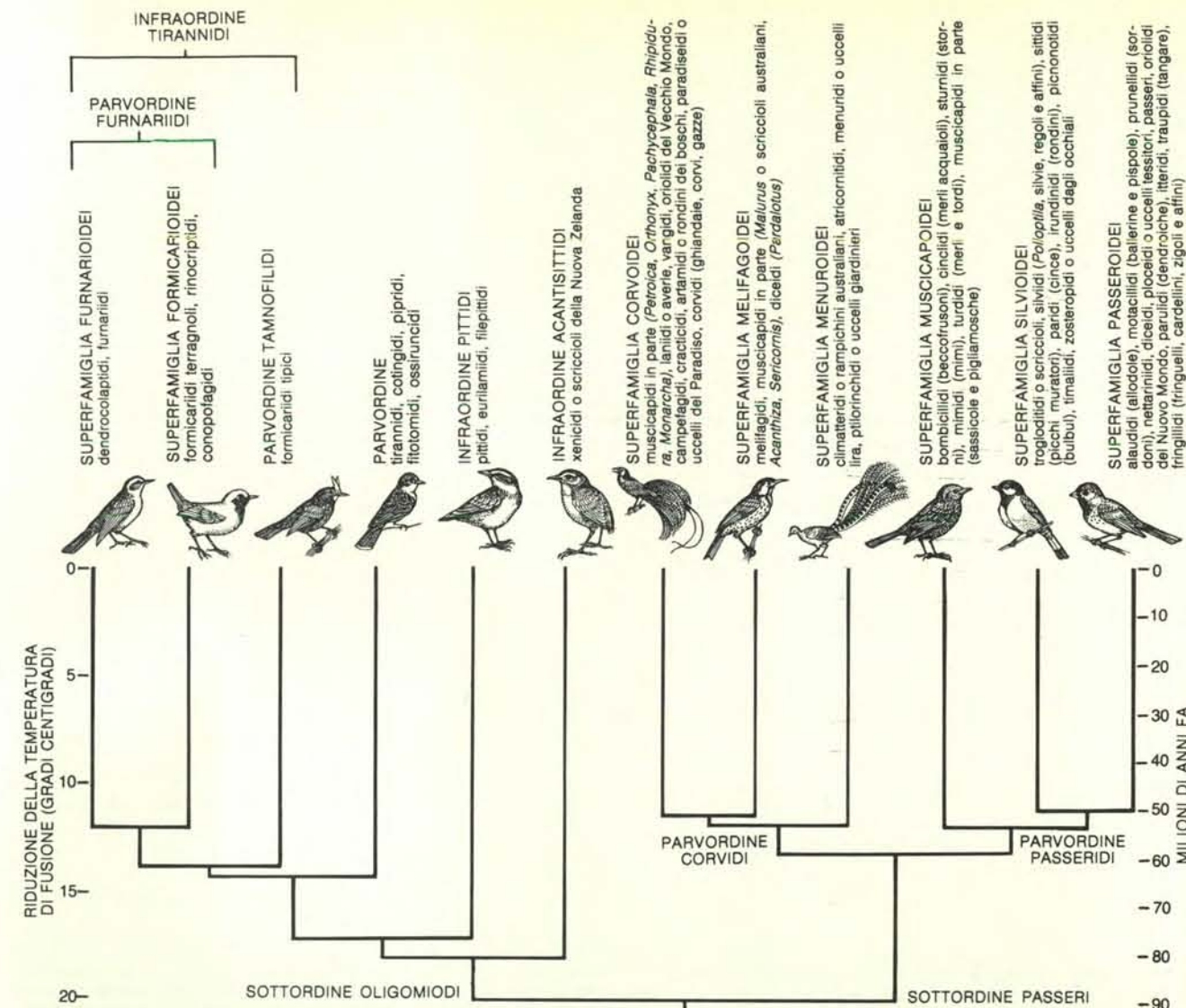
Da quanto abbiamo potuto verificare, sembra che l'orologio del DNA scandisca i tempi, in media, con lo stesso ritmo in tutte le linee filetiche degli uccelli: la procedura che abbiamo seguito, proposta nel 1967 da Vincent M. Sarich e Allan C. Wilson, dell'Università della California a Berkeley, è nota come test della velocità relativa. Questo tipo di analisi serve a confrontare tra loro tre specie qualsiasi, due delle quali risultano più strettamente imparentate di quanto ciascuna delle due lo sia con la terza. Se si scelgono tre specie di uccelli con queste caratteristiche e si mette a confronto il DNA tracciante della specie filogeneticamente più distante con il DNA elemento guida di ciascuna delle altre due specie, risulta che le distanze genetiche tra la specie più distante e ciascuna delle altre due

Con il test della velocità relativa si può verificare se la velocità media di evoluzione dei DNA di specie differenti è o no la stessa. Il test può essere condotto su tre specie qualsiasi, due delle quali (*B*, *C*) siano più strettamente imparentate di quanto ciascuna di esse lo sia con la terza (*A*). Il DNA delle prime due specie viene ibridato con quello della terza: se le temperature di fusione dei due ibridi risultano identiche, ciò significa che essi contengono un egual numero di accoppiamenti errati di basi. Quindi, le distanze genetiche tra le specie *A* e *B* (*in colore*), e tra le specie *A* e *C* (*in grigio*) sono le stesse. Poiché lo stesso lasso di tempo separa le specie *B* e *C* dall'ultimo antenato condiviso con la specie *A*, le velocità medie di mutazione dei loro DNA devono essere identiche.

specie, indicate dalle temperature di fusione degli ibridi, sono costantemente identiche, ovviamente nei limiti dell'errore sperimentale. Poiché lo stesso arco di storia evolutiva separa entrambe le specie che hanno fornito il DNA elemento guida dall'ultimo antenato in comune con la specie che ha fornito il DNA tracciante, il DNA delle prime due specie deve essere mutato alla stessa velocità media. Nel nostro lavoro di ricostruzione della filogenesi degli uccelli viventi, abbiamo incluso centinaia di queste «triplette» di specie che hanno prodotto tutte gli stessi risultati dimostrando così l'uniformità del ritmo con cui l'orologio molecolare, costituito dal DNA degli uccelli, scandisce i tempi.

L'analizzatore di DNA, cioè l'apparecchio che permette di analizzare i DNA ibridi, opera simultaneamente su 25 DNA ibridi. Le operazioni sono sincronizzate da un temporizzatore, mentre un termostato regola la temperatura dell'acqua in cui sono immerse a bagnomaria le 25 colonne di idrossiapatite contenenti i campioni di DNA ibrido. Quando il temporizzatore dà il segnale d'avvio, un nastro trasportatore porta i flaconi in posizione, sotto le colonne; le valvole al fondo di ciascuna colonna si aprono e consentono così alla soluzione tampone, fornita da una pompa peristaltica, di fluire attraverso le colonne. La soluzione eluisce dalle colonne, dentro flaconi, i filamenti di DNA singoli, prodotti dalla fusione degli ibridi. Le valvole quindi si richiudono, e la temperatura dell'acqua nella vasca che contiene le colonne viene innalzata di 2,5 gradi centigradi; il temporizzatore emette un nuovo segnale e il ciclo ricomincia. Una stima della quantità di DNA marcato, che è contenuto nei vari flaconi, indica quanta parte di ciascun DNA ibrido si è fusa a ogni temperatura.

Questa apparente costanza potrebbe sembrare magica (o senza senso), ma potrebbe anche essere la conseguenza del fatto che si misurano differenze tra sequenze composte da miliardi di coppie di basi dopo milioni di anni di evoluzione. La selezione naturale fa sì che geni diversi si evolvano a velocità varie e che un gene qualsiasi si evolva a velocità diverse in tempi diversi. Tuttavia, l'intervallo entro cui variano le velocità di evoluzione di tutti i geni è piccolo e il numero di geni nel genoma di un uccello è enorme. Mentre la velocità di evoluzione di un gene aumenta, è statisticamente probabile che quella di un altro gene diminuisca in ugual misura. Il bilanciamento tra le velocità di mutazione dei vari geni non deve essere necessariamente simultaneo poiché l'apparente costanza insorge dopo milioni di anni. Qualunque possa essere la spiegazione corretta, il DNA degli uccelli sembra evolversi con una velocità media uniforme.

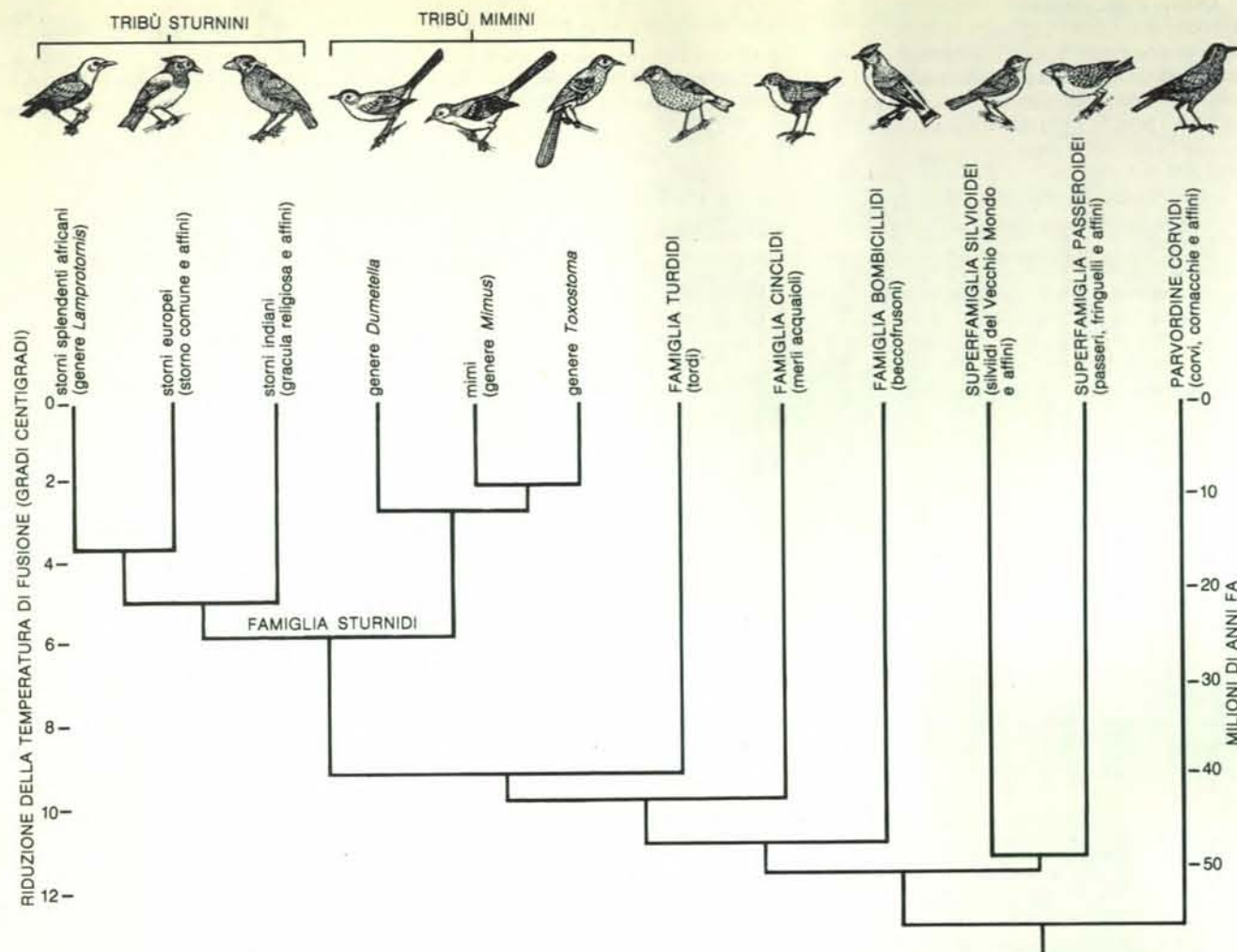
L'errato appaiamento dei DNA ibridi risulta da cambiamenti genetici che si sono fissati nelle due linee filetiche dal momento in cui, per l'ultima volta, hanno avuto un antenato in comune: pertanto il numero di appaiamenti errati è proporzionale al tempo trascorso dalla divergenza delle due linee. La temperatura di fusione media di un ibrido DNA-DNA è quindi una misura indiretta di questo tempo. Si può tarare in tempo assoluto l'orologio molecolare costituito dal DNA, mettendo in relazione la sua temperatura di fusione media con un dato evento geologico, che ha causato la separazione di due diverse linee filetiche da una stessa specie ancestrale.

Per esempio, abbiamo supposto che l'antenato comune dello struzzo africano e del nandù sudamericano vivesse nel Gondwana prima che la deriva dei continenti, nel Cretaceo, fratturasse il Gondwana negli attuali continenti dell'emisfero australe dando origine all'oceano Atlantico. I dati geologici indicano che l'Atlantico divenne una barriera geografica per gli organismi incapaci di volare circa 80 milioni di anni fa: le linee filetiche dello struzzo e del nandù, quindi, incominciarono a divergere in quel periodo. Dividendo 80 milioni di anni per la differenza tra la temperatura di fusione media dell'eteroduplex DNA di struzzo/DNA di nandù e la temperatura di fusione media dell'omoduplex DNA di struzzo/DNA di struzzo o DNA di nandù/DNA di nandù si ricava la costante di taratura, espressa in milioni di anni di divergenza per grado centigrado di riduzione della temperatura di fusione media.

Conosciamo tre casi di analoghe divergenze datate tra linee filogenetiche di uccelli, causate da tre diversi eventi geologici, due prossimi agli 80 milioni di anni fa e uno che risale a circa 40 milioni di anni fa. In ciascuno di essi, la costante di taratura era compresa tra 4,3 e 4,7, con un valore medio di 4,5. Quindi una riduzione media di un grado centigrado nella temperatura di fusione equivale a un tempo di circa 4,5 milioni di anni che è trascorso da quando le due linee filetiche hanno avuto per l'ultima volta un antenato comune. Questa costante è ancora provvisoria e suscettibile di correzioni; tuttavia la usiamo ugualmente per una datazione approssimativa delle divergenze.

La filogenesi dei passeriformi, che comprendono 5300 delle 9000 specie viventi di uccelli, è stata ricostruita tramite confronti di DNA. La scala a sinistra indica, per ciascuna biforcazione dell'albero, di quanti gradi si è abbassato il punto medio di fusione degli ibridi formati con i DNA di specie rappresentative delle due linee filetiche rispetto a DNA ibridi di controllo, cioè che abbiano un perfetto accoppiamento delle basi. La scala a destra indica invece i tempi nei quali si sono avute le divergenze. Questa ricostruzione mette in discussione la filogenesi tradizionale degli uccelli viventi: nel sottordine oligomiodi, al quale appartiene la maggior parte dei passeriformi sudamericani, i formicariidi terragnoli risultano appartenenti a una linea filetica diversa da quella dei formicariidi tipici. Nel sottordine passeri si delineano due gruppi ben distinti: i parvordini corvidi e passeridi. I primi hanno avuto origine in Australia, anche se alcune famiglie sono oggi distribuite nelle varie parti del mondo.

Negli ultimi 10 anni abbiamo eseguito più di 25 000 confronti di ibridi DNA-DNA, usando il materiale genetico di circa 1600 specie, che nell'insieme costituiscono 168 delle 171 famiglie tradizionali in cui sono classificati gli uccelli viventi. Su questi dati, abbiamo ricostruito la filogenesi della maggior parte dei gruppi di uccelli: gli esempi che seguono illustrano alcuni dei problemi esaminati e le soluzioni indicateci dai dati sui DNA ibridi.

I barbuti sono uccelli piccoli e dai colori generalmente brillanti, caratterizzati da ciuffi di setole alla base dei becchi relativamente grossi. Le specie del Vecchio Mondo vivono in Africa e nell'Asia meridionale, mentre quelle del Nuovo Mondo vivono nelle regioni tropicali dell'America Centrale e Meridionale. Tradizionalmente, essi vengono considerati membri della famiglia capitonidi e ritenuti affini ai picchi (picidi) e ai tucani (ranfastidi), questi ultimi uccelli frugivori dai grandi becchi, diffusi nell'America tropicale. Molti sistematici hanno notato somiglianze morfologiche considerevoli tra le piccole specie di tucani e le grosse specie di barbuti del Nuovo Mondo, mentre due studi recenti hanno messo in rilievo la stretta relazione tra tucani e barbuti. Nel 1984, Philip Burton del British Museum of Natural History ha così concluso uno studio anatomico sulla regione cefalica: «Sembra ragionevole considerare i tucani come un gruppo specializzato di barbuti che hanno avuto origine e si sono diffusi nell'America Meridionale». Analogamente, nel 1985, Lester L. Short, Jr., dell'American Museum of Natural History ha affermato: «I tucani sono effettivamente grossi e specializzati barbuti dal becco dentellato».

I confronti tra i DNA danno risultati concordi a Burton e Short, e ci consentono anche di aggiungere la dimensione temporale alle loro conclusioni. Il DNA indica che la divergenza tra i barbuti del Vecchio e del Nuovo Mondo avvenne circa 55 milioni di anni fa; i tucani, invece, si separarono dalla linea dei barbuti del Nuovo

In base ai dati di ibridazione DNA-DNA, sull'albero filogenetico degli uccelli le tribù degli storni (sturnini) e dei mimi (mimini) occupano rami adiacenti. Gli sturnini sono stati considerati affini ai corvi; i mimini ai tordi. Se ciò fosse esatto, la divergenza tra i due gruppi dovrebbe risalire a 60 milioni di anni fa. Ma le temperature di fusione dei DNA ibridi indicano che sturnini e mimini sono tra loro i parenti più stretti, con una divergenza non oltre i 25 milioni di anni fa. Entrambi i gruppi che costituiscono la famiglia sturnidi sono poi imparentati con i tordi (turdidi).

Mondo circa 30 milioni di anni fa. Quindi, i tucani sono affini ai barbuti del Nuovo Mondo più di quanto lo siano i due gruppi di barbuti tra loro. La classificazione da noi proposta per questi gruppi e per gli altri membri del medesimo ordine, tenendo conto di queste relazioni, è la seguente:

ordine piciformi
 parvordine picidi
  famiglia picidi (picchi)
  famiglia indicatoridi (indicatori)
 parvordine ranfastidi
  superfamiglia megalaimoidei
   famiglia megalaimidi (barbuti del Vecchio Mondo)
  superfamiglia ranfastoidei
   famiglia ranfastidi
    sottofamiglia ranfastini (tucani)
    sottofamiglia capitonini (barbuti del Nuovo Mondo)

La nostra classificazione è fondata sullo schema di ramificazione dell'albero filogenetico e sui tempi in cui hanno avuto origine i vari gruppi, datati con confronti tra DNA. Abbiamo suddiviso la scala del tempo in segmenti di 10 milioni di anni e abbiamo assegnato una categoria tassonomica a ogni segmento. Gli ordini sono quelle linee filetiche che si separarono da altre linee 90-100 milioni di anni fa, mentre i sottordini sono quelle linee che si separarono 80-90 milioni di anni fa. Gli infraordini ebbero origine 70-80 milioni di anni fa, i parvordini 60-70 milioni di anni fa, le superfamiglie 50-60 milioni di anni fa, le famiglie 40-50 milioni di anni fa, le sottofamiglie 30-40 milioni di anni fa e le tribù 20-30 milioni di anni fa. Uno dei vantaggi di questo modello è che gruppi sistematici dello stesso rango sono circa equivalenti nel grado di divergenza evolutiva: lo schema che abbiamo costruito non è definitivo, ma ha il pregio di fornire i fondamenti per una classificazione che rifletta la filogenesi basata sul DNA e contemporaneamente si avvicini all'ideale di equivalenza dei gruppi sistematici.

Tuttavia non c'è ancora accordo su come classificare gli organismi. Alcuni studiosi preferiscono valutare il rango di un gruppo sistematico in base alla stima del suo grado di specializzazione morfologica: in questa ottica, i due gruppi di barbuti verrebbero inclusi nella stessa famiglia, mentre i tucani finirebbero in una famiglia o superfamiglia adiacente. Una classificazione siffatta rileva la diversità d'aspetto degli uccelli, ma nasconde le loro relazioni genealogiche. Nello schema tradizionale, ancora largamente usato, tucani, barbuti e altri gruppi affini sono classificati così:

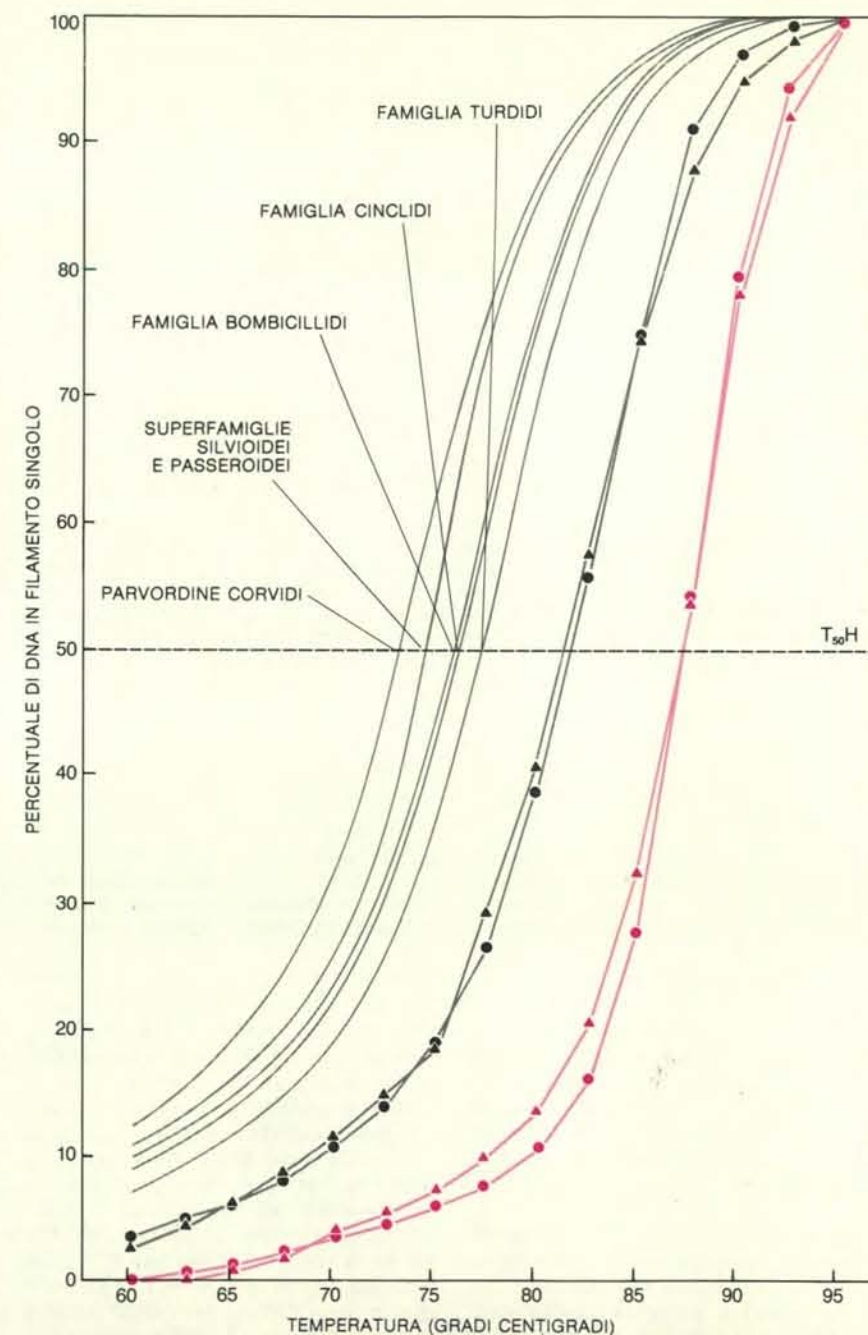
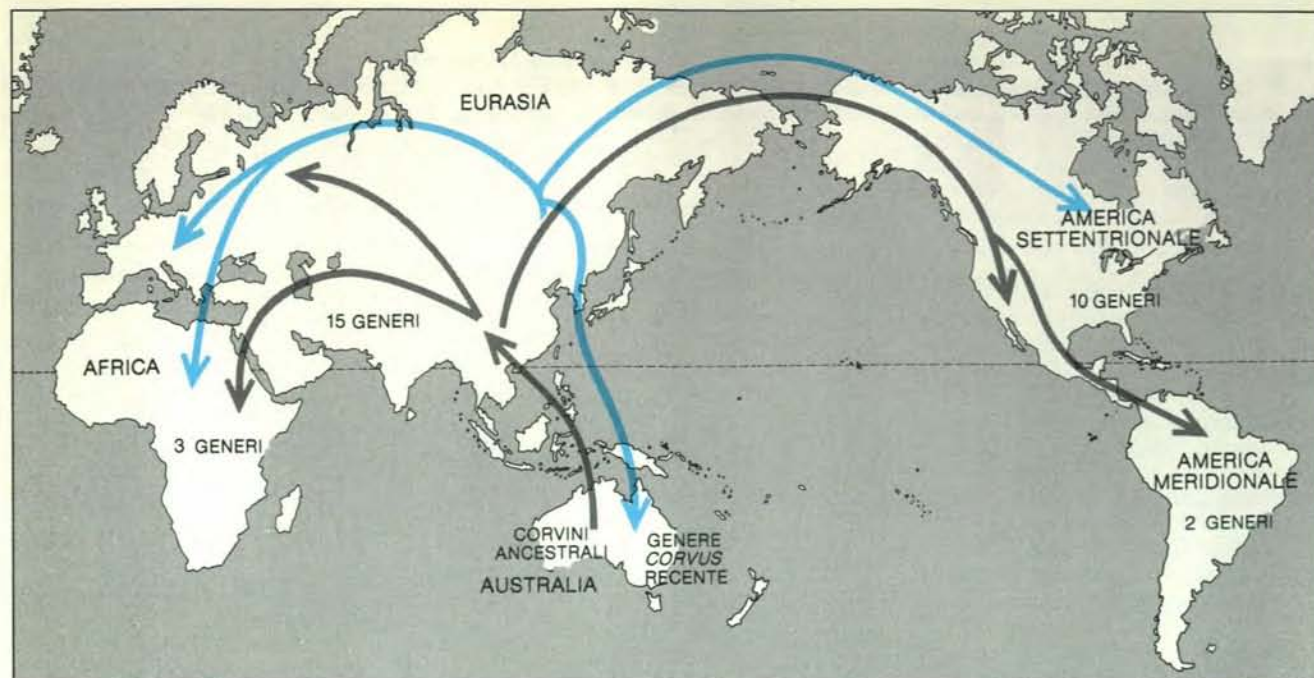
ordine piciformi
 sottordine galbuli
  superfamiglia galbuloidei
   famiglia galbulidi (jacamar)
   famiglia bucconidi (bucchi)
  superfamiglia capitonoidei
   famiglia capitonidi (barbuti)
   famiglia indicatoridi (indicatori)
  superfamiglia ranfastoidei
   famiglia ranfastidi (tucani)
 sottordine pici
   famiglia picidi (picchi)

Questa classificazione è in netto contrasto con i dati sul DNA, non soltanto riguardo alla posizione di tucani e barbuti, ma anche relativamente a jacamar e a bucchi del Nuovo Mondo, che in base a essi andrebbero posti in un ordine separato, e non nei piciformi. Per il momento sembra proprio che non vi sia alcun segno di rapida conclusione del dibattito sul modo di impostare la classificazione delle piante e degli animali.

Gli avvoltoi del Vecchio Mondo sono strettamente imparentati con aquile e falchi; quelli del Nuovo Mondo, che comprendono i condor, l'avvoltoio dal collo rosso e l'avvoltoio nero o gallinazo, hanno solo una somiglianza superficiale con i primi; entrambi i gruppi, infatti, si nutrono di carogne e sono sempre stati inclusi assieme nell'ordine falconiformi, in cui sono compresi tutti gli uccelli predatori diurni. Gli avvoltoi del Nuovo Mondo condividono, tuttavia, molti caratteri morfologici con le cicogne e alcuni sistematici hanno sostenuto che i due gruppi appartengono al medesimo ordine. Questo ordinamento è stato proposto fin dal decennio 1870-1879 da Alfred B. Garrod e, nel 1967, da David Ligon, allora all'Università del Michigan. Una serie di recenti classificazioni ha però ignorato i pur evidenti legami tra i condor e le cicogne e ha continuato a mettere i condor e gli uccelli loro affini nell'ordine falconiformi.

I confronti tra i DNA confermano i risultati delle indagini morfologiche e suggeriscono che avvoltoi del Nuovo Mondo e cicogne siano, gli uni nei confronti degli altri, i parenti più vicini tra gli esseri viventi: i dati sul DNA mostrano, infatti, che i due gruppi iniziarono a divergere da un antenato comune in un periodo compreso tra 35 e 40 milioni di anni fa. Le somiglianze esteriori tra gli avvoltoi del Nuovo e del Vecchio Mondo sono dovute a un'evoluzione convergente, connessa con le loro abitudini di divorare carogne.

Vi è un gruppo di uccelli, detti «totipalmi» o «totipalmati», che è caratterizzato dall'avere tutte e quattro le dita unite da membrana; comprende i pellicani, i cormorani, le aninghe, le sule, le fregate o aquile di mare e gli uccelli del sole o fetontidi. Tutti, tranne gli uccelli del sole, hanno una voluminosa sacca golare, situata fra i due rami della mandibola; negli uccelli del sole la sacca è, invece, piccola e nascosta. A causa di questi due caratteri comuni, gli uccelli totipalmati sono sempre stati raggruppati insieme come membri dell'ordine pelecaniformi, anche se talvolta si è suggerito che uccelli del sole e fregate siano affini a qualche altro gruppo.

Un altro uccello, il becco a scarpa (*Balaeniceps rex*), è stato recentemente proposto come membro dello stesso raggruppamento: simile a una cicogna e con un enorme becco, esso vive nelle paludi dell'Africa orientale, nutrendosi di dipnoi e di altre prede acquatiche. Il becco a scarpa è stato generalmente considerato affine alle cicogne e agli aironi, ma nel 1957, studiandone lo scheletro, Patricia A. Cottam, allora al British Museum, ha stabilito che è vicinissimo ai pellicani. Le conclusioni della Cottam sono state respinte dalla maggior parte dei sistematici, che hanno considerato le somiglianze con i pellicani come il risultato di convergenze evolutive. I nostri dati sul DNA, le danno, invece, ragione: la divergenza tra il becco a scarpa e i pellicani ebbe luogo tra 35 e 40 milioni

Le curve di fusione dei DNA ibridi rivelano l'entità delle differenze tra il DNA degli sturnini (storni) e quello dei mimini (mimi) e tra il DNA di uno sturnino o di un mimino e quello di altre linee filetiche degli uccelli. La scala orizzontale mostra le temperature alle quali gli ibridi di DNA sono stati riscaldati; la scala verticale indica, invece, la percentuale di DNA che fondeva a ciascuna temperatura, dando origine a filamenti singoli. Le curve in colore indicano le proprietà di fusione di DNA ibridi, i cui filamenti provengono entrambi da *Toxostoma longirostra*, una specie affine ai mimi (*pallini*), o da *Lamprotornis*, un genere di storni splendenti africani (*triangoli*). Le curve in nero, immediatamente a sinistra delle precedenti, mostrano l'andamento della fusione di DNA ibridi di mimini con varie specie di sturnini (*pallini*) o di sturnini con varie specie di mimini (*triangoli*). Le altre curve indicano le proprietà medie di fusione di DNA ibridi, composti da un filamento di DNA di mimino, o di sturnino, e da un filamento di DNA di una specie appartenente al gruppo indicato per ogni curva. Per confrontare gli ibridi si ricorre alla temperatura a cui metà di un ibrido si è fusa ($T_{50}H$). Quanto più bassa è la $T_{50}H$ di un DNA ibrido, tanto più scarso è il grado di appaiamento tra le basi dei due filamenti e tanto più distanti sono tra loro le specie interessate.

La dispersione della tribù corvini, comprendente corvi, cornacchie, ghiandaie, gazze e loro affini, è indicata sulla cartina dalle frecce in grigio. L'ibridazione DNA-DNA ha identificato i corvini come parte di un più vasto gruppo di uccelli, evolutisi in Australia. Il numero di generi di corvini presenti in ciascun continente indica l'ordine con cui i continenti stessi sono stati occupati dai loro antenati quando, circa 30 milioni di anni fa, passarono dall'Australia in Asia. Il maggior numero di generi è presente in Asia e il più piccolo nell'America Meridionale, regione che i corvini raggiunsero solo 5-3 milioni di anni fa. Corvi e cornacchie (genere *Corvus*) ebbero origine in Asia e si diffusero in seguito in Africa, in Europa e nell'America Settentrionale (*frecce in colore*). Solo di recente, all'incirca negli ultimi 100 000 anni, il genere ha colonizzato l'Australia.

di anni fa, mentre il raggruppamento più vicino al gruppo dei pellicani e dei becco a scarpa risulta essere quello che comprende, insieme, avvoltoi del Nuovo Mondo e cicogne. La divergenza dei due gruppi ebbe inizio da 40 a 45 milioni di anni fa.

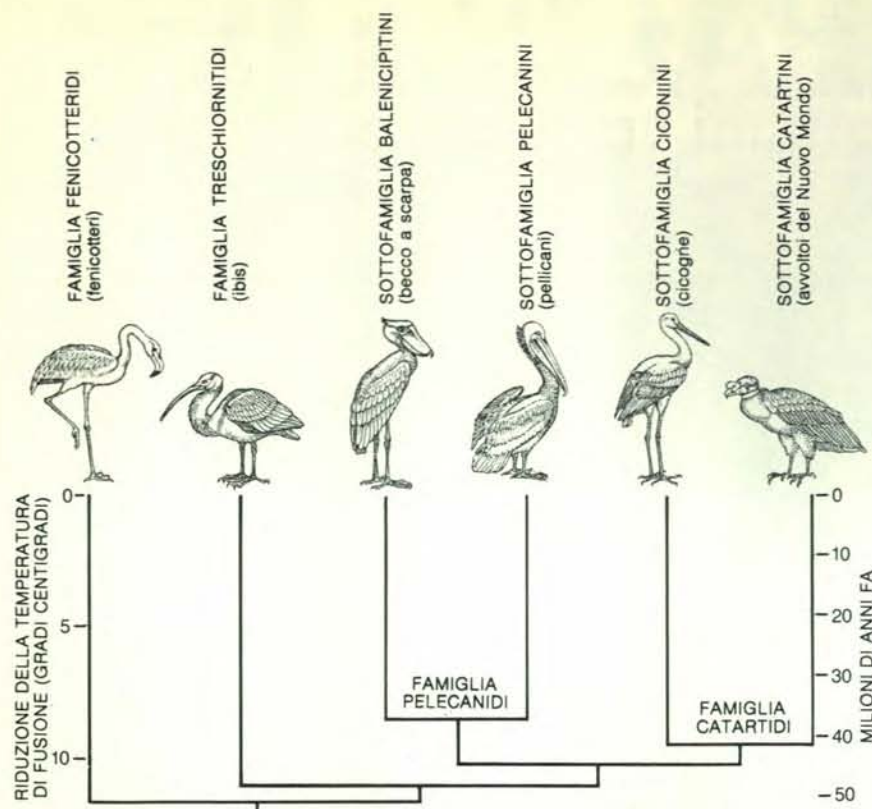
E le sorprese non sono ancora finite: le ricerche sul DNA dimostrano anche che, in realtà, il tradizionale ordine pelecaniformi è un gruppo polifiletico. Infatti, esso è composto da numerosi sottogruppi, che sono imparentati con altri gruppi più strettamente di quanto non lo siano tra loro: i pellicani sono fortemente legati ai becco a scarpa, alle cicogne e ai condor; cormorani, aninghe e sule costituiscono un unico gruppo lontanamente imparentato con gli aironi; le fregate sono affini ai procellariformi (albatros e uccello delle tempeste); gli uccelli del sole, infine, sembrano appartenere a una diversa linea evolutiva e non hanno affini tra i gruppi di uccelli viventi. Quindi, sia la zampa totipalmata sia la sacca golare potrebbero essere caratteri affermatisi per convergenza evolutiva in linee filetiche distinte; o, in alternativa, potrebbero essere state ereditate da un lontano antenato, comune a molti diversi gruppi di uccelli. Nei discendenti che non presentano tali caratteri, i geni a essi relativi possono essere stati repressi. Geni silenti di questo tipo sono noti in alcuni gruppi di uccelli e in altri animali.

L'origine polifiletica di alcuni gruppi di uccelli era già stata riconosciuta in passato: in alcune delle prime classificazioni, gli uccelli con zampe palmate (cioè con solo le tre dita frontali unite da membrana) erano raggruppati insieme. Ci volle però poco a capire che non tutti gli uccelli dalle zampe palmate - anatre, albatros, pinguini, strolaghe, gabbiani e alche - erano tra loro affini ed essi vennero perciò assegnati a gruppi diversi. Invece, nel caso degli uccelli totipalmati (tutte e quattro le dita unite da membrana), accomunati anche dall'avere una sacca golare, sembrava impossibile che essi non fossero un raggruppamento monofiletico, cioè un raggruppamento i cui componenti hanno tutti in comune un antenato recente.

In uno studio di non molto tempo fa, Joel Cracraft del Medical Center dell'Università dell'Illinois a Chicago ha messo a confronto gli uccelli totipalmati, sulla base di 45 caratteri scheletrici e di sette caratteri comportamentali, con pinguini, strolaghe, svassi, albatros, uccelli delle tempeste e becco a scarpa. L'ipotesi suggerita da Cracraft, che gli uccelli totipalmati siano un gruppo monofiletico, risulta avvalorata dalla comunanza di 12 caratteri, mentre sei caratteri lasciano supporre che albatros e uccelli delle tempeste costituiscano un gruppo (o taxon) fratello (*sister group* secondo Hennig: si veda in proposito l'articolo *Il systema naturae* di Alessandro Minelli in «Le Scienze» n. 206, ottobre 1985). Gli uccelli del sole, sempre secondo Cracraft, emersero come linea filetica distinta in mezzo al gruppo di tutti gli altri totipalmati. Di questi uccelli, le fregate sono i discendenti del ramo più antico, seguite in ordine successivo di divergenza dai pellicani, dalle sule, dalle aninghe e dai cormorani.

Cracraft rifiuta di considerare il becco a scarpa come affine ai totipalmati e attribuisce a convergenza la sua somiglianza con i pellicani. Vi sono quindi evidenti discordanze tra i risultati ottenuti confrontando i DNA e i dati morfologici che Cracraft ha utilizzato. Molte conclusioni sono però le stesse: la relazione degli uccelli totipalmati (in particolare le fregate) con albatros e uccelli delle tempeste, la posizione del tutto separata degli uccelli del sole e le strette affinità tra sule, aninghe e cormorani. Le discordanze principali riguardano il collegamento tra i becco a scarpa e i pellicani o la posizione dei pellicani rispetto agli altri uccelli totipalmati.

È improbabile che la maggior parte degli ornitologi sia disposta ad accettare subito la natura polifiletica dei pelecaniformi, così come risulta dallo studio del DNA. Tuttavia, prevediamo che confronti adeguati metteranno in luce affinità morfologiche che, come lo studio della Cottam, nel 1957, siano coerenti con i dati sul DNA.

Gli pteroclidi (sirratti e grandule) sono uccelli simili nell'aspetto ai colombi o ai pivieri che vivono nelle regioni aride dell'Africa, dell'Asia e dell'Europa meridionale. Le loro relazioni con gli altri uccelli sono oggetto di discussione da più di un secolo: gli pteroclidi sono imparentati con i colombi, con i pivieri o con i galliformi

Secondo la filogenesi ricostruita in base ai confronti tra DNA, avvoltoi del Nuovo Mondo e cicogne sono, gli uni rispetto agli altri, i parenti viventi più vicini. Superficialmente, gli avvoltoi del Nuovo Mondo (che comprendono tra l'altro i condor) assomigliano agli avvoltoi del Vecchio Mondo: queste somiglianze però sono dovute a convergenza evolutiva, perché gli avvoltoi del Vecchio Mondo appartengono a un gruppo diverso di linee filetiche. Gli studi sui DNA mostrano anche che i pellicani e i becco a scarpa, una specie africana, sono il gruppo più vicino a condor e cicogne nell'albero filogenetico. Il DNA ha confermato la stretta parentela tra becco a scarpa e pellicani.

(galli, fagiani e via dicendo)? Ogni gruppo ha avuto i suoi sostenitori e la maggior parte dei più recenti partecipanti al dibattito propende per i piccioni o per i pivieri.

I dati sui DNA sono inequivocabili: gli pteroclidi costituiscono il gruppo fratello di una grande parte dell'ordine caradriformi, che comprende chioni, occhioni, pivieri, beccacce di mare (ostrichieri), avocette, cavalieri, gabbiani, alche e corrioni. Gli pteroclidi e i loro affini sono a loro volta i più stretti affini di un gruppo che comprende i piro-piro e simili. Quindi, essi non sono né colombi né pivieri, ma sono più affini ai pivieri che ai colombi; le somiglianze morfologiche con i colombi sono dovute a un'evoluzione convergente.

Circa 5300 specie di uccelli, fra le 9000 viventi, appartengono all'ordine passeriformi, che comprende pigliamosche, silvie, tordi, passeri, storni, scriccioli, rondini, allodole, corvi e molte altre specie, generalmente di piccole dimensioni. La maggior parte dei passeriformi sudamericani è costituita da membri del sottordine oligomiodi, noti anche come suboscini. I suboscini sono distinti dagli oscini o uccelli canori (sottordine passeri, l'altro ramo dei passeriformi) per la struttura della siringe (apparato vocale) e per altri caratteri anatomici. I suboscini del Nuovo Mondo si sono evoluti nell'America Meridionale durante il periodo (iniziato 80 milioni di anni fa - nell'ultima fase del Cretaceo - e terminato 5 milioni di anni fa), in cui quella parte del continente americano era rimasta isolata dal resto del mondo.

Tra i suboscini del Nuovo Mondo è presente un gruppo dalla interessante storia evolutiva, i formicariidi: circa 235 specie tropicali tradizionalmente riunite in quest'unica famiglia. All'inizio degli anni sessanta, Mary Heimerdinger Clench e Peter L. Ames, entrambi alla Yale University, scoprirono che alcuni membri della famiglia avevano due profonde intaccature nel margine posteriore dello sterno, mentre altre specie ne avevano quattro. Ames trovò anche che i due gruppi differivano per la muscolatura della siringe. Le 185 specie con due intaccature occupano una varietà di habitat, mentre le 50 specie con quattro intaccature vivono al suolo e sono tutte caratterizzate dall'avere lunghe zampe e coda corta. La Clench e Ames notarono anche che altri due gruppi di uccelli possiedono uno sterno con quattro intaccature: i rinocriptidi e i conopofagidi. Secondo loro i formicariidi terragnoli sono più vicini a questi due gruppi che non agli altri formicariidi.

Le ricerche sul DNA hanno confermato del tutto quest'ipotesi: i confronti hanno messo in luce, infatti, che i formicariidi con due intaccature si separarono dalla linea filetica di quelli a quattro intaccature prima che questi divergessero dai rinocriptidi e dai conopofagidi. Così i dati morfologici e quelli molecolari risultano concordanti e i risultati delle ricerche sul DNA forniscono l'ordine delle ramificazioni successive e i tempi approssimativi in cui si sono realizzate le divergenze.

L'altro ramo dei passeriformi, il sottordine passeri, comprende circa 4000 delle 5300 specie del gruppo. I confronti tra i DNA rivelano che esso si suddivide, a sua volta, in due grandi gruppi che abbiamo chiamato parvordini: i corvidi e i passeridi. Queste due linee filetiche hanno iniziato a divergere da un antenato comune 55-60 milioni di anni fa. Si hanno dati secondo i quali i passeridi si sarebbero evoluti in Africa, in Europa e nell'America Settentrionale e i corvidi in Australia.

Dall'inizio a metà del Terziario (circa 60-30 milioni di anni fa), l'Australia è rimasta isolata dalle altre masse continentali e i corvidi, nella loro evoluzione, hanno dato origine a molte forme ecologicamente e morfologicamente specializzate, comprendenti silvie, pigliamosche, rampichini, tordi, timaliidi, nettarinie - forme molto simili a quelle che i passeridi svilupparono in altre parti del mondo. Tuttavia, questi uccelli australiani vennero scoperti e descritti quando gli ornitologi europei avevano ormai classificato quasi tutti gli altri uccelli del mondo; i passeri australiani sembrarono però inserirsi bene nelle categorie che erano già state istituite per gli altri passeriformi. Così, quelli simili alle silvie e alle capinere furono assegnati ai silviidi (che comprendono le vere silvie), i pigliamosche australiani ai muscicapidi (i pigliamosche afroeuroasiatici), e i rampichini australiani ai certiidi (rampichini euroasiatici e americani). Le sittelle, simili ai picchi muratori, vennero messe tra i sittidi (la famiglia dei veri picchi muratori) e i melifagidi australiani vennero raggruppati con i nettariniidi afroasiatici, superficialmente simili.

Quando abbiamo confrontato i DNA dei vari passeri australiani tra loro e delle specie australiane con quelli dei loro ipotetici affini africani, euroasiatici e nordamericani, abbiamo trovato che le forme endemiche australiane erano tra loro molto più vicine di quanto lo fossero nei confronti delle rispettive controparti, morfologicamente simili e viventi in altri continenti. Un'evoluzione convergente aveva prodotto analogie tra specie non affini dei due parvordini e i tassonomisti avevano riunito queste specie in gruppi polifiletici contenenti membri sia dei corvidi sia dei passeridi. Lo stesso errore era stato ripetuto per la maggior parte delle 400 specie di passeri dell'Australia e della Nuova Guinea. Parecchie convergenze riscontrate sono così sottili che le vere relazioni tra corvidi e passeridi non sarebbero state svelate, con tutta probabilità, unicamente da

confronti tra i vari caratteri anatomici.

Una delle conseguenze di questa confusione è stata quella di occultare per molto tempo l'affascinante storia della filogenesi dei corvidi, del tutto parallela a quella dei marsupiali. Entrambi i gruppi, infatti, si sono evoluti mentre l'Australia era isolata; come i passeri australiani, anche i marsupiali si sono irradiati in molte delle nicchie occupate in Europa, in Asia, in Africa e in Nordamerica da un altro gruppo, in questo caso i mammiferi placentati, e alcuni di essi hanno assunto, durante questo processo, forme simili a quelle di altri mammiferi. Diversamente dai passeri australiani, però, i marsupiali non furono confusi con le loro controparti di altri continenti perché erano contraddistinti dal marsupio e da altri caratteri anatomici.

Tra i corvidi australiani si annoverano gli antenati di alcuni gruppi che migrarono poi in Asia quando, nel Terziario, l'Australia andò alla deriva verso nord: uno dei gruppi i cui antenati ebbero origine in Australia è la tribù corvini, che comprende le stesse specie della famiglia tradizionale corvidi, e cioè cornacchie, corvi, ghiandaie, gazze e affini.

La tribù corvini conta oggi 23 generi: 15 di questi vivono in Eurasia e 10 nell'America Settentrionale. Solo tre generi sono stati trovati in Africa, mentre nell'America Meridionale la tribù è rappresentata soltanto da due generi di ghiandaie. Queste cifre svelano la cronologia e la modalità di dispersione dall'Australia: la prima radiazione avvenne probabilmente nell'Asia sudorientale e successivamente i membri delle varie linee evolutive estesero i loro areali fino a comprendere l'Europa, l'Africa e l'America Settentrionale. L'America Meridionale è stato l'ultimo continente a essere invaso, anche perché esso è rimasto isolato dall'America Settentrionale fino a 3-5 milioni di anni fa: i due generi di ghiandaie che si trovano oggi nell'America Meridionale sono probabilmente arrivati dall'America Settentrionale, estendendo i loro areali quando l'emersione dell'America Centrale ha fornito un collegamento terrestre tra i due continenti. Corvi e cornacchie del genere *Corvus* hanno avuto probabilmente origine in Europa e di lì si sono irradiati sostanzialmente ovunque (eccetto che nell'America Meridionale), ricolonizzando alla fine anche l'Australia.

Tra i molti sorprendenti risultati riservatici dai confronti tra DNA, nessuno è stato così inaspettato come l'aver scoperto la stretta relazione di parentela che esiste tra altri due gruppi di passeri: gli storni, che sono originari del Vecchio Mondo, e i mimi, del continente americano. Gli storni sono stati considerati, in genere, affini ai corvi, mentre i mimi venivano correttamente posti vicino ai tordi. Se gli storni fossero davvero parenti stretti dei corvi, essi sarebbero membri dei corvidi; d'altra parte, il DNA identifica chiaramente i mimi come membri dei passeridi. Quindi, se la classificazione tradizionale fosse corretta, le due linee filetiche avrebbero dovuto divergere tra 55 e 60 milioni di anni fa.

I confronti tra i DNA hanno rivelato, invece, una divergenza tra le due linee avvenuta circa 25 milioni di anni fa. Altri studi convalidano l'esistenza di questa stretta relazione tra storni e mimi: confronti immunologici tra proteine muscolari, condotti nel 1961 da William B. Stallcup, Jr., della Southern Methodist University, studi anatomici della regione cefalica, condotti nel 1953 da William J. Beecher, allora all'Università di Chicago, e confronti riguardanti la struttura della siringe, effettuati da Wesley E. Lanyon dell'American Museum of Natural History. Può anche essere significativo il fatto che alcuni storni (per esempio, la gracula religiosa), sono, al pari dei mimi, eccellenti imitatori. Anche lo storno comune (*Sturnus vulgaris*), del resto, imita il canto di altri uccelli.

La stretta relazione tra storni e mimi può essere il riflesso della storia dei mutamenti climatici nell'emisfero boreale; infatti, all'inizio e alla metà del Terziario (65-30 milioni di anni fa), le regioni artiche godevano di un clima di tipo temperato e foreste di latifoglie crescevano nel Canada settentrionale e in Groenlandia. È probabile, quindi, che il comune antenato di storni e mimi fosse ampiamente distribuito in queste regioni, che costituivano un ponte tra il Vecchio e il Nuovo Mondo. A quanto si può ricostruire dalle piante fossili, l'inasprimento climatico iniziò circa 30 milioni di anni fa, inducendo le popolazioni ancestrali a migrare verso sud: infine, circa 25 milioni di anni fa, esse si separarono nei continenti americano ed euroasiatico e iniziarono a divergere.

Queste sono solo alcune delle scoperte rese possibili dalla tecnica di ibridazione DNA-DNA. I confronti tra DNA ci offrono nuove ipotesi sulle relazioni filogenetiche tra gli uccelli: se la filogenesi così ricostruita rispecchia veramente la situazione reale essa dovrebbe ricevere conferma anche da altre fonti di dati. I risultati che abbiamo finora ottenuto sono coerenti con la storia geologica più di quanto lo fossero molte precedenti proposte sull'origine e sull'evoluzione degli uccelli. Inoltre, di solito, i dati riguardanti il DNA trovano conferma in almeno alcuni caratteri anatomici. Noi siamo convinti che alcuni aspetti morfologici si dimostreranno sempre in accordo, in tutti i casi considerati, con i dati che emergono dal DNA.

Così, per ricostruire la filogenesi, il DNA da una parte e i caratteri morfologici usati tradizionalmente dall'altra permettono di raccogliere diversi tipi di informazione: i caratteri morfologici dimostrano in che modo la selezione naturale abbia modellato le strutture per adattare gli organismi all'ambiente, mentre il confronto tra vari DNA dà un'indicazione diretta dello schema di ramificazione dell'albero filogenetico e dei tempi approssimativi di divergenza tra le varie linee evolutive oggi viventi. La morfologia diventa così funzionale e l'orologio del DNA ne scandisce i tempi.

(RI)CREAZIONI AL CALCOLATORE

di A. K. Dewdney

Il re (un programma per gli scacchi) è morto. Viva il re (una macchina per gli scacchi)!

Se CRAY BLITZ fosse in grado di ricordare qualcos'altro oltre le mosse degli scacchi, non dimenticherebbe mai la sera del 15 ottobre 1985. In quella sera si è tenuta l'ultima tornata del Campionato nordamericano di scacchi per calcolatori, in occasione del convegno annuale dell'Association for Computing Machinery. Cinque tavoli, separati dal pubblico da una barriera, sono sistemati all'inizio di una sala del Radisson Hotel di Denver. A ciascun tavolo si fronteggiano due squadre di programmatori ed esperti che un po' giocano, un po' rimangono in posizione di attesa.

Il torneo prevede la partecipazione di 10 contendenti al titolo, dai nomi alquanto bizzarri: AWIT, BEBE, CHAOS, CRAY BLITZ, HITECH, INTELLIGENT SOFTWARE, LACHEX, OSTRICH, PHOENIX e SPOC (*si veda l'illustrazione della pagina a fronte*). Sono assenti tre grandi nomi che hanno dominato gli scacchi al calcolatore negli anni scorsi: BELLE, CHESS 4.7 e NUCHESS.

L'interesse è incentrato sulla partita finale tra CRAY BLITZ e HITECH. Al tavolo, dalla parte di CRAY BLITZ, stanno Robert Hyatt dell'Università del Southern Mississippi, Albert Gower, un esperto di scacchi dello stesso ateneo, e Harry Nelson del Lawrence Livermore National Laboratory. I loro avversari sono Hans Berliner della Carnegie Mellon University e Murray Campbell, uno dei suoi studenti che è un giocatore esperto. Nella squadra di HITECH, Berliner ricopre il duplice ruolo di programmatore e di esperto. Col procedere della gara e il crescere della tensione, Berliner si alza spesso, con un sorriso stanco sul volto. Una volta passa vicino alla mia sedia e mi sussurra: «È proprio come ai tempi dei miei campionati nazionali!» (Per parecchi anni, tra la fine degli anni cinquanta e l'inizio degli anni sessanta, Berliner si è qualificato tra i primi dodici giocatori degli Stati Uniti.)

A differenza di quel che succede ai campionati degli Stati Uniti, dove regna un silenzio mortale, in questo torneo si conversa, talvolta si ride, si sente il rumore continuo delle tastiere e l'incessante commento dell'arbitro Michael Valvo, un eccentrico consulente di informatica e maestro internazionale di scacchi di Sedona, Arizona. «Mossa debole da parte del nero. Il re è ancora troppo esposto e i due pedoni in c5 e c6 continuano a ostacolare la difesa.» Lì accanto un menbro della squadra di CRAY BLITZ esclama, senza rivolgersi a qualcuno in particolare; «Divertente! Pensavo che avrebbe messo il re in f3».

Durante tutta questa tornata finale è risultato chiaro che HITECH ha il sopravvento sul suo rivale: ben presto CRAY BLITZ è caduto in un *zugzwang*, una posizione critica da cui si può uscire solo con una brutta mossa o con una perdita di pezzi. In questo caso, CRAY BLITZ è costretto a disporre male i suoi pedoni e HITECH continua a sfruttare il vantaggio.

A mezzanotte non è ancora finita. La maggior parte delle partite sono terminate e gli esperti affermano che HITECH ha vinto. La squadra di CRAY BLITZ chiede al giudice Valvo il permesso di abbandonare e Valvo propone di fare altre due mosse: se la posizione di CRAY BLITZ non migliora, la squadra può abbandonare. Così avviene: HITECH è campione nordamericano e di fatto re degli scacchi al calcolatore. Anche se CRAY BLITZ rimane ufficialmente campione del mondo (ha vinto il titolo nel 1983 e non deve difenderlo fino a giugno), il successo di HITECH, insieme alle sue vittorie in altri tre tornei, è impressionante. Quasi certamente, è il più forte calcolatore per scacchi del mondo.

Si ride e si continua a conversare. L'assenza di BELLE, CHESS 4.7 e NUCHESS è significativa? Dice un organizzatore del torneo: «Sarebbe stato interessante se BELLE e qualche altro programma avessero partecipato, ma non credo che il risultato sarebbe stato molto diverso». Prosegue precisando che, in termini di programmi e di macchine, sostanzialmente non c'è differenza tra campionato nordamericano e campionato del mondo. Il discorso si sposta su Kasparov e Karpov e poi sulla teoria. Dice un partecipante, evidentemente degno di considerazione: «Non sto scherzando. Un programma a 20 livelli che si preoccupi solo dei pezzi può battere qualsiasi gran maestro». L'affermazione suscita un po' di discussione, ma dopo qualche minuto la sala si svuota: il campionato nordamericano è terminato.

La battuta sul programma a 20 livelli è interessante. Il gioco degli scacchi si può rappresentare come un grande albero fatto di linee e nodi. Lo rappresento rovesciato cosicché il nodo radice si trovi in alto. Ciascun nodo rappresenta una posizione possibile, vale a dire una scacchiera sulla quale i pezzi e i pedoni siano arrivati alle loro posizioni attraverso mosse legittime. Un nodo è congiunto da una linea a un nodo discendente, se lo spostamento di un singolo pezzo o di un pedone trasforma la scacchiera rappresentata dal primo nodo in quella rappresentata dal secondo. Una partita di scacchi si può sempre identificare con un particolare cammino lungo l'albero degli scacchi, a partire dal nodo radice (quando non è stata fatta alcuna mossa) giù lungo l'albero fino a un nodo dove, come regola generale, sono rimasti pochi pezzi e uno dei giocatori ha subito scacco matto o è stato costretto ad abbandonare.

Un programma che gioca a scacchi cerca di esplorare, dell'albero, solo quel tanto che serve. Dal nodo che rappresenta la posizione attuale esamina tutte le scacchiere discendenti (livello 1), poi le discendenti delle discendenti (livello 2) e così via. La profondità media della sua esplorazione viene detta «previsione». In questa misura è compresa la maggior parte di quella che si potrebbe chiamare l'intelligenza del programma. In parte minore, essa consiste nella valutazione delle scacchiere che costituiscono l'orizzonte della sua previsione. Il programma le analizza e attribuisce a ciascuna un valore numerico che ne riflette l'appetibilità. Con una procedura chiamata minimax, il programma fa risalire alcuni valori lungo l'albero fino ai nodi del livello 1. Il nodo che riceve il valore più alto indica il gioco da svolgere.

Vi è un'interessante complementarità tra le due parti dell'intelligenza del programma: tanto migliore è lo schema di valutazione, tanto meno profonda deve essere l'esplorazione dell'albero. In effetti, se il programma avesse uno schema di valutazione perfetto, non avrebbe bisogno di esplorare più a fondo del primo livello. Al contrario, un programma con uno schema di valutazione molto semplice deve indagare molto più in profondità, se vuole giocare in maniera efficace. A quale profondità deve scendere una ricerca che si preoccupi solo dei pezzi, per essere efficace contro un gran maestro? È sufficiente una ricerca di venti livelli?

Il titolo di gran maestro è conferito dalla Fédération Internationale des Echecs a giocatori che si sono distinti in campo internazionale. (La federazione non prende in considerazione i calcolatori.) Un gran maestro ha generalmente punteggi superiori a 2400, il livello di un maestro anziano. Fino al Campionato nordamericano di scacchi per calcolatori, HITECH aveva giocato 21 partite in tornei umani, guadagnando 2233 punti. Era, cioè, il più quotato fra i calcolatori che giocano a scacchi in tutto il mondo. Secondo Berliner, che ai tempi in cui gareggiava aveva un punteggio di 2443, la quotazione di HITECH è cresciuta con una media di otto punti a partita nei tornei internazionali. Si può supporre che, dopo altre 14 partite, la macchina sorpasserà il suo ideatore?

Tutto questo porta a porsi la domanda: quanto bravi possono diventare i calcolatori che giocano a scacchi? Un calcolatore diventerà mai il migliore giocatore del mondo? David Levy, un tempo giocatore e attualmente autore e imprenditore, ha affidato la questione a una serie di scommesse. Nel 1968 scommise 500 sterline con John McCarthy della Stanford University che nessun calcolatore lo avrebbe battuto nei successivi 10 anni. Levy incassò la posta nell'agosto del 1978 in occasione della Canadian National Exhibition di Toronto, dove giocò con CHESS 4.5, un programma elaborato alla Northwestern University. La scommessa iniziale venne, poi, rinnovata per un periodo di altri sei anni e per 6000 dollari. Nell'aprile 1984 Levy da Londra ha giocato per telefono con CRAY BLITZ e ha vinto di nuovo.

Tutto questo lo ha spinto a scommettere, a Denver, 100 000 sterline che nei prossimi 10 anni qualunque calcolatore che raccolga la sfida verrà battuto da un giocatore umano scelto da lui. Se qualcuno raccoglierà la sfida di Levy, sarà probabilmente non un semplice programma, ma un calcolatore specializzato. Finora non si è fatto avanti alcuno sfidante.

I due finalisti del torneo nordamericano, HITECH e BEBE, erano proprio macchine di questo tipo. Curiosamente, il concorrente dello stesso Levy, un programma chiamato INTELLIGENT SOFTWARE, si è classificato terzo. Gira su un Apple IIe che non ha nulla di più sofisticato di una scheda acceleratrice, che ne raddoppia la velocità. Forse Levy ha sviluppato uno schema di valutazione superiore.

Gli esperti di scacchi presenti al campionato sono d'accordo nel ritenere che la miglior partita del torneo sia stata quella della seconda tornata tra CRAY BLITZ e BEBE, un prodotto dell'impresa privata. Tony Scherzer, la cui azienda, SYS-10 Inc., ha sviluppato BEBE, negli ultimi anni ha portato la sua creazione a parecchi tornei. BEBE non è un semplice programma, ma una vera e propria macchina che gioca a scacchi. La partita è stata significativa non solo perché era la più interessante del torneo, ma anche perché era la prima partita persa da CRAY BLITZ in tre anni.

I lettori con una scacchiera possono seguire la partita di CRAY BLITZ contro BEBE giocando le 50 mosse elencate nel seguito. I pezzi sono indicati da lettere maiuscole: R, re; D, regina; A, alfiere; C, cavallo; T, torre. Le caselle della scacchiera sono contrassegnate dalle coordinate lettere-numeri. Quando la scacchiera è nella posizione standard, cioè con la casella in basso a sinistra nera, le colonne sono contrassegnate da sinistra a destra con le lettere da *a* fino a *h*; le traverse sono numerate da 1 a 8 a cominciare dal basso della scacchiera. La notazione impiegata nella presentazione delle partite può andare dalla semplice formula Rb1 (re nella casella b1) alla problematica Cf3 (cavallo in f3). Quale cavallo? In quella particolare mossa solo un cavallo può spostarsi in f3. La mossa di un pedone viene indicata designando una casella, per esempio e4.

La partita è commentata da Valvo.

	CRAY BLITZ (Bianco)	BEBE (Nero)
1.	e4	c5
2.	Cf3	d6
3.	d4	cxd4

(La x significa che un pezzo o un pedone viene mangiato.)

4.	Cxd4	Cf6
5.	Cc3	g6
6.	Ag5	Ag7
7.	Dd2	Cc6
8.	0-0-0	0-0

(Arrocco lungo del bianco, arrocco corto del nero.)

9.	Cb3	Te8
10.	Ac4	Cg4

Il nero ha giocato Cg4 pensando di fare Ac3xC alla mossa successiva. Il nero può aver pensato il bianco obbligato a 12 c3xA, ma il nero cambia parere alla successiva mossa del bianco. Se Axc3, allora Dxc3!; Cxf2 non funziona se l'una o l'altra delle due torri viene portata in f1, mentre Axf7, scacco, sarebbe fatale.

11.	h3	Cge5
12.	Ab5	a6
13.	Ae2	a5
14.	Ab5	Ae6
15.	Cd5	a4
16.	Cd4	Ad7

La situazione del bianco è disperata. Il pedone nero in a4 minaccia di creare una difficile situazione di debolezza intorno al re bianco.

17.	Cxc6	bxc6
18.	Cxe7 scacco	Txe7
19.	Axe7	Dxe7
20.	Ae2	De6
21.	Rb1	Tb8

(Nell'illustrazione di pagina 106 si può vedere la situazione della scacchiera a questo punto.) Il nero minaccia 22...Tb2 scacco, seguito da 23 Rb2 Cc4, una forchetta che elimina la regina bianca.

22.	b3	axb3
23.	cxb3	Ae8
24.	Rc2	Cd7
25.	f3	Ta8
26.	Rc1	Cc5
27.	Dc2	Df6
28.	Ac4	Da1 scacco
29.	Rd2	Dxa2

Una mossa ancora più forte è 29...Ac3, scacco, seguita da 30 Re2 Ta2!

30.	Dxa2	Txa2 scacco
31.	Rc1	d5
32.	exd5	cxd5
33.	Axd5	Ab5
34.	The1	Cd3 scacco

PROGRAMMA	ORIGINE	CALCOLATORE	LINGUAGGIO	POSIZIONI AL SECONDO	LIVELLI DI PREVISIONE
AWIT	Università dell'Alberta	Amdahl 5860	Algol W	10	3
BEBE (secondo)	SYS-10, Inc., Hoffman Estates, Illinois	Macchina personalizzata	Assembler	20 000	7
CHAOS	Università del Michigan	Amdahl 5860	FORTRAN	70	4
CRAY BLITZ	Università del Southern Mississippi	Cray X-MP 48	FORTRAM/ Assembler	100 000	8
HITECH (primo)	Carnegie-Mellon University	Sun con circuiti VLSI personalizzati	C	175 000	8
INTELLIGENT SOFTWARE (terzo)	Intelligent Soft-Ware Inc., Londra	Apple II e con scheda acceleratrice	Assembler	500	7
LACHEX	Los Alamos National Laboratory	Cray X-MP 48	FORTRAN/ Assembler	50 000	7
OSTRICH	McGill University	Rete di sette Nova e un Eclipse	Assembler	1200	6
PHOENIX	Università dell'Alberta	Rete di VAX 780 e 10 Sun	C	540	6
SPOC	SDI/Cypress Software, San Jose, California	IBM PC	Assembler	300	5

I partecipanti al Campionato nordamericano di scacchi per calcolatore

Il vantaggio (un pezzo) del nero sta per essere aumentato da un altro scambio. In un torneo tra uomini a questo punto sarebbe ragionevole per il bianco abbandonare.

35.	Txd3	Axd3
36.	Te8 scacco	Af8
37.	g4	Rg7
38.	Te3	Aa3 scacco
39.	Rd1	Ta1 scacco
40.	Rd2	Af1
41.	Rc3	Tc1 scacco
42.	Rd2	Tc5
43.	Re1	Axh3
44.	Ac4	h5
45.	gxh5	gxh5
46.	Rf2	h4
47.	Td3	Af5
48.	Td4	h3
49.	Th4	Tc7
50.	Th5 (chiede di abbandonare)	

Il programma CRAY BLITZ gira su un calcolatore Cray XM-P 48. Famoso per la sua velocità come multielaboratore, il Cray è nondimeno un calcolatore di uso generale e non una macchina per gli scacchi. BEBE, i cui circuiti sono dedicati agli scacchi, ha ovviamente superato, nella partita riportata qui sopra, la combinazione Cray-CRAY BLITZ.

HITECH in un certo senso è più specializzato. Quando la Carnegie-Mellon University era il Carnegie Institute of Technology, vi venne messo a punto un programma che giocava a scacchi, chiamato TECH. Il nome HITECH è legato al fatto che Berliner, Campbell e gli altri membri del gruppo HITECH (Carl Ebeling, Gordon Goetsch, Andy Palay e Larry Slomer) hanno fatto rivivere la tradizione di TECH in un mondo di integrazione a grandissima scala (VLSI: *very large scale integration*) e di crescente parallelismo. La macchina HITECH comprende un calcolatore Sun dotato di un eleboratore appositamente progettato, che Berliner chiama il ricercatore. Sul calcolatore Sun girano tre programmi: un'interfaccia utente, un controllore di compito e un oracolo. L'oracolo comprende quello che gli esperti di scacchi con il calcolatore chiamano il libro, ossia un grande catalogo di aperture e variazioni che fanno parte del bagaglio comune dei giocatori esperti di scacchi. La base di dati dell'oracolo contiene molte altre nozioni scacchistiche, che si possono facilmente espandere. Quando il ricercatore esamina le possibilità di gioco a partire da una posizione data, procede sulla base delle informazioni rilevanti rispetto a quella posizione ricavate dall'oracolo.

Il ricercatore stesso contiene un microelaboratore e parecchi moduli hardware che generano mosse, le valutano, controllano le mosse ripetute e così via. Il microelaboratore coordina la loro attività. Il generatore di mosse è fatto di 64 chip VLSI, uno per ciascuna casella della scacchiera. Ciascun chip esamina l'intera scacchiera per stabilire se un pezzo o un pedone possono essere portati sulla casella sotto il suo controllo. Esso determina la mossa migliore secondo criteri standard come quelli che stabiliscono quando è opportuno catturare un pezzo o controllare il centro. Nello stesso momento gli altri 63 chip stanno facendo la stessa cosa. Questa struttura comporta che, se ci sono 10 pezzi sulla scacchiera, le mosse possibili vengono generate dieci volte più velocemente, mantenendo invariati gli altri fattori.

La valutazione delle mosse deve procedere di pari passo con il processo di generazione delle mosse stesse. Una prima fase di valutazione viene compiuta dal generatore stesso, che ospita una specie di supervisore, il quale giudica le mosse generate dai 64 chip. Ciascun chip calcola un numero che dà una valutazione della potenza della sua mossa migliore e trasmette il numero al supervisore. I numeri generati dai chip sono come urla che chiedono attenzione. Il supervisore li dispone in ordine di volume (leggi: di efficacia).

HITECH procede, quindi, ad analizzare l'albero di gioco secondo l'ordinamento prodotto dal supervisore delle mosse che sono possibili a partire dalla posizione di quel momento. Una seconda fase di valutazione è condotta dal modulo di valutazione in ciascuna nuova posizione che si genera all'interno dell'albero di gioco. Servendosi delle nozioni rilevanti rispetto alla posizione attuale, fornite dall'oracolo, il modulo valuta ciascuna disposizione sulla scacchiera, sia che si trovi, sia che non si trovi nell'orizzonte di previsione. Il parallelismo consiste in questo. Un maggior impegno non comporta un maggior tempo. Il controllore di compito del Sun dice al ricercatore quanto andare a fondo nell'esplorare l'albero e, quando la ricerca è terminata, se procedere ancora oltre. In questo modo HITECH gestisce una previsione media di otto livelli, ma può, se se ne presenta l'occasione, esplorare anche 14 livelli. Questo può sembrare ancora lontano dai 20 livelli necessari per battere un gran maestro. L'uso fatto da HITECH del parallelismo, però, e il raffinato impiego di nozioni scacchistiche nell'esplorazione dell'albero compensano forse la relativamente scarsa profondità dell'indagine. In ogni caso, al momento del prossimo Campionato del mondo di scacchi per calcolatori, che si terrà a giugno, a Colonia, nella Germania Federale, HITECH sarà probabilmente inarrestabile.

HITECH è la più recente macchina per scacchi del mondo. La prima fu inventata nel 1890 da Leonardo Torres y Quevedo, un ingegnere spagnolo. Servendosi di leve meccaniche, pulegge e interruttori elettromeccanici, giocò una mediocre partita di torre e re contro re. Agli uomini venne concesso il privilegio di curare gli affari di un re solo, cercando di evitare di subire scacco matto da parte della potente com-

La scacchiera dopo la ventunesima mossa

binazione di torre e re della macchina. La macchina di Torres y Quevedo non fu mai sconfitta.

I lettori dovrebbero creare una strategia che produca questo risultato. Si deve supporre che il re del giocatore uomo non inizi in una posizione di stallo. Si tratta, quindi, di precisare nel minor numero di regole possibile come fa la macchina a dare scacco matto a partire da una posizione arbitraria. La posizione illustrata nella figura di questa pagina può costituire un punto di partenza

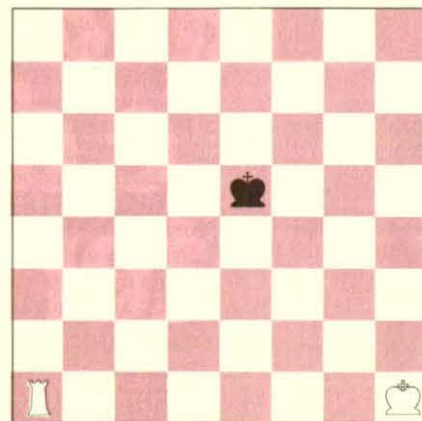
La macchina è il bianco e muove per prima. Come fanno il re bianco e la torre a costringere il re nero allo scacco matto? Il bianco potrebbe cominciare col muovere la sua torre nella traversa *d*. Questo impedirebbe al nero di spostare il suo re a sinistra. La manovra può essere ripetuta, se il re nero si sposta verso destra, ma che cosa succede se continua a occupare la traversa *e* andando semplicemente avanti e indietro? Pubblicherò la soluzione più rapida, indipendentemente dal fatto che venga espressa in linguaggio algoritmico, o in linguaggio naturale.

Come dare scaccomatto con torre e re?

Le «(Ri)creazioni al calcolatore» di gennaio parlavano dei flib (*finite living blobs*): bolle viventi finite che cercano di prevedere i cambiamenti del loro ambiente. Nel brodo primordiale del calcolatore, durante ogni generazione il miglior previsore incrocia i cromosomi con un flib a caso. Si evolvono previsori sempre più precisi, finché ne emerge uno perfetto.

Un flib è sostanzialmente un automa finito, cioè ha un numero finito di stati e, per ogni segnale che riceve (uno 0 o un 1), invia un segnale ed entra in un nuovo stato. Il segnale emesso dal flib a ogni ciclo di attività è quello che prevede il successivo segnale ricevuto dall'ambiente.

Alcuni lettori hanno dato ai loro flib dei compiti di previsione impossibili. Non si evolverà mai un flib che possa prevedere una successione di bit casuale e neppure i numeri primi, mentre è perfettamente ragionevole chiedere a un flib di prevedere una successione binaria periodica. Per esempio, c'è un flib a quattro stati che prevederà la successione periodica a otto simboli 01100010. Persino una successione periodica, tuttavia, mette alla prova la capacità di previsione di un flib, se la sua successione base è troppo lunga rispetto al numero di stati del flib. Nessun flib a quattro stati prevederà mai la successione periodica 010010111. Perché no?

La risposta più semplice comporta un processo che io chiamo induzione strisciante. Si immagini un flib a uno stato che potrebbe prevedere la ripetizione infinita della successione base 01. Per ciascuno dei due possibili segnali che il flib riceve c'è una sola risposta: se il flib riceve uno 0, manda un 1 e poi ritorna nello stesso stato. Se riceve un 1, manda uno 0. Una successione base di tre simboli, per esempio 011, è al di fuori della capacità di previsione del flib a uno stato, semplicemente perché l'automa non ha un repertorio sufficiente di risposte. Un flib a due stati, d'altra parte, ha quattro possibili risposte, due per ogni stato. Può perciò prevedere una successione periodica di quattro simboli; quando si arriva a cinque simboli, il flib deve ripetere una risposta precedente. Il ragionamento è chiaro. Un flib a n stati può prevedere una successione base lunga $2n$ simboli, ma non una lunga $2n + 1$ simboli. È una piacevole distrazione ideare una successione base di otto simboli e poi costruire con carta e matita il flib che la prevederà. Il previsore perfetto ottenuto è sostanzialmente unico. Si può misurare il successo di un programma AUTOSOUP paragonando il previsore perfetto che da esso si evolve con il flib già costruito.

Parecchi lettori hanno trovato il modo per far girare più velocemente AUTOSOUP. Per esempio, non vale la pena di mettere alla prova l'attuale insieme di flib su una successione di 100 simboli ambientali, se la successione base è lunga solo sei simboli. Una ripetizione della successione produrrà 12 simboli ambientali, che dovrebbero bastare nella maggior parte dei casi.

Philip Kaaret della Princeton University ha precisato che il programma si può anche abbreviare se, a ogni esecuzione del ciclo principale, si considerano due flib invece dell'intera popolazione. Dopo tutto, due soli flib (al più) sono cambiati: il flib con i punteggi peggiori è stato sostituito da un nuovo ibrido e forse un altro flib è stato colpito da un raggio cosmico.

Gli aumenti di velocità che si ottengono, rispettivamente abbreviando la sequenza di controllo ambientale ed eliminando il controllo per i vecchi flib, sono pressappoco equivalenti. Ora ci sarà tempo per sviluppare flib a n stati che possono prevedere successioni base ripetute lunghe fino a $2n$ simboli.

Dalla sua lettera si capisce che Ed Coudal di Park Ridge, Illinois, era riluttante a spedire direttamente al Creatore il suo flib con il peggior punteggio. Lo ha invece incrociato a ogni ciclo con il flib col punteggio più alto. Seguendo questo schema, in meno di 40 generazioni Coudal ha potuto ottenere flib capaci di prevedere una successione base di sei simboli.